

Robust designs for longitudinal studies

Citation for published version (APA):

Ortega Azurduy, S. A. (2009). *Robust designs for longitudinal studies*. [Doctoral Thesis, Maastricht University]. Datawyse / Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.20090701so>

Document status and date:

Published: 01/01/2009

DOI:

[10.26481/dis.20090701so](https://doi.org/10.26481/dis.20090701so)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Robust designs for longitudinal studies

Shirley Amparo Ortega Azurduy

In memoriam Nelly y Jorge

Robust designs for longitudinal studies

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Universiteit Maastricht, op gezag van de
Rector Magnificus, prof. mr. G.P.M.F. Mols, volgens het besluit
van de College van Decanen, in het openbaar te verdedigen
op woensdag 1 juli 2009 om 14.00 uur

door

Shirley Amparo Ortega Azurduy



Promotor:

Prof. dr. M.P.F. Berger

Copromotor:

Dr. E.S. Tan

Beoordelingscommissie/Evaluation committee:

Prof. dr. M.H. Prins (voorzitter)

Dr. G.J.P. van Breukelen

Dr. ir. M. Moerbeek (Universiteit Utrecht)

Prof. dr. G. Molenberghs (Universiteit Hasselt)

Prof. dr. ir. R.L.M. Peeters

Copyright © 2009 by Shirley Amparo Ortega Azurduy, Maastricht
Universitaire Pers Maastricht, 2009.

ISBN 978-90-5278-845-6

This dissertation was supported by the Netherlands Organisation for Scientific Research (NWO, Grant number: 400 – 03 – 100) and the School for Public Health and Primary Care (Caphri) of the Faculty of Health, Medicine and Life Sciences (FHML) of Maastricht University.

Contents

1	Introduction	9
1.1	An example of longitudinal study	10
1.2	Theoretical background	13
1.2.1	The design	13
1.2.2	The linear mixed model and its assumptions	14
1.2.3	Probability of response	15
1.2.4	Asymptotic variance-covariance of estimators	16
1.2.5	Optimality criteria	17
1.2.6	A manner to investigate robustness of designs	19
1.3	Robustness of designs	21
1.3.1	Robustness of D -optimal designs against the incorrect specification of the degree of the polynomial model	22
1.3.2	Robustness of D -optimal designs against the incorrect specification of the variance-covariance matrix of the random effects	22
1.3.3	Robustness of D -optimal designs against the incorrect specification of the error variance-covariance structure	23
1.3.4	Robustness of D -optimal designs against the dropout	24
1.3.5	Robustness of A -optimal designs against incorrect assumptions on weights and design scales	24
1.4	The goal	25
1.5	Organization of the thesis	25

2	Highly efficient designs to handle the incorrect specification of linear mixed models	27
2.1	Introduction	29
2.2	Motivating example	30
2.3	Linear mixed models and optimality of designs	31
2.3.1	Optimal designs	32
2.3.2	Maximin criterion	34
2.4	Numerical analysis	37
2.4.1	Algorithm	39
2.5	Results	41
2.6	Special cases	44
2.6.1	Discussion on Hamilton Depression Scores'example . .	45
2.7	Conclusions	46
3	On the D-optimal designs of misspecified heterogeneous linear mixed models	49
3.1	Introduction	51
3.2	Longitudinal study of elderly chronic care populations	52
3.3	Linear mixed model	52
3.3.1	Optimal designs	54
3.3.2	Measure of efficiency change	55
3.4	Classes of linear mixed models	55
3.4.1	Homogeneous and heterogeneous error covariance structure	57
3.5	Results	57
3.5.1	D -optimal designs of random intercept models	57
3.5.2	Efficiency loss	58
3.5.3	Consequences for the longitudinal study of elderly chronic care populations	63
3.6	Conclusions	64
4	The effect of dropout on the efficiency of D-optimal designs of linear mixed models	67
4.1	Introduction	69
4.2	Illustration: Randomized clinical trial	70
4.3	Linear mixed models	71
4.3.1	Design and response probability function	72
4.3.2	Linear mixed models with dropout	73
4.3.3	D -Optimal designs	74
4.3.4	Relative efficiency	75

4.4	Numerical analysis	76
4.4.1	Dropout function	77
4.4.2	Classes of linear mixed models	77
4.4.3	Algorithm	78
4.4.4	D -optimal designs for data with dropout	78
4.4.5	Efficiency loss due to the incorrect assumption of complete data	79
4.4.6	D -optimal designs and efficiency loss of the randomized clinical trial example	84
4.5	Conclusions	86
5	On the scale dependency of A-optimal designs for random intercept models with serially correlated errors	89
5.1	Introduction	91
5.2	Random-intercept model with serially correlated errors	92
5.2.1	Design, A -optimality criterion and relative efficiency .	93
5.3	Numerical method	95
5.4	Results	96
5.4.1	A -Optimal designs	96
5.4.2	Relative efficiency of designs applied to different intervals	100
5.4.3	Relative efficiency of designs applied under equal weight restriction	104
5.5	Discussion and conclusion	106
6	ODMixed: A tool to obtain optimal designs for heterogeneous longitudinal studies with dropout	109
6.1	Introduction	111
6.2	Motivating examples	112
6.2.1	Example 1: Randomized clinical trial on psycho- and pharmacotherapy	112
6.2.2	Example 2: Longitudinal study of elderly chronic care populations	113
6.3	Longitudinal data and linear mixed models	113
6.3.1	Optimal designs and response probability function . .	114
6.3.2	Asymptotic covariance matrix	115
6.3.3	Generalized variance of the estimated parameters . . .	116
6.3.4	Relative efficiency to compare designs	117
6.3.5	Algorithm	117
6.4	ODMIXED	118

6.4.1	Example 1 (Revisited): Randomized clinical trial on psycho- and pharmacotherapy	121
6.4.2	Example 2 (Revisited): Longitudinal study of demen- tia populations	124
6.5	Discussion and conclusions	127
References		130
7	Summary and recommendations	139
7.1	Guidelines for researchers	141
7.2	Further research	143
8	Samenvatting in het Nederlands (Dutch Summary)	145

CHAPTER 1

Introduction

In past years countless longitudinal studies have been conducted in biomedical, clinical and epidemiological research. Longitudinal studies are a general purpose approach to characterize changes in response variable(s) over time and to examine covariates which contribute to that change (Everitt, 1998). As a result, processes can be described, understood and/or predicted. Practitioners, for example, may apply the outcomes of such longitudinal studies to exert a positive influence on human health. The pattern description of cancer growth, the development of mental diseases caused by substance abuse or the prediction of the efficacy of a dementia treatment, are just a few of the many examples of longitudinal studies that one can think of (Goldstein, 1979; Taris, 2000; Verbeke and Molenberghs, 2000; Ishak et al., 2007).

Researchers must decide whether either an observational study or an experimental study best suits their research goals. *Observational studies* are distinguished because research conditions are not all under the researcher's control. Subjects are divided into "natural" or intrinsic categories and are followed-up in time. Typical examples of these studies are prospective cohorts, retrospective cohorts and case-control studies. *Experimental studies* are characterized by controlled research conditions. Subjects are "exposed" to different levels of the treatment and the allocation of subjects to each group is usually at random. Randomized clinical trials (RCT) belong to this category. For more details on longitudinal designs see e.g., Taris (2000), Rosenbaum (2002), Polit and Beck (2004) and Bouter et al. (2005).

A design is longitudinal when the collection of data will occur at different points in time. Thus the researcher plans a systematically tracking of change in response over time prior to data collection. As a result, the responses can be modelled as a function of time. When additional factors (variables) are appropriate for modelling a specific process, this variable is then also tracked and added to the model. A distinction must be made between *within-* and *between-subject* factors (variables). In longitudinal designs the focus is on the variable time, which is always a within-subject factor, i.e. data are time-structured.

Researchers prefer designs which are parsimonious and intuitively appropriate for their specific research problems. An example of such a design is the model-free equally-spaced design where the time points are distributed at fixed and evenly distributed points regardless of any model or model assumptions. It has been shown however, that this is not the best strategy (Tan and Berger, 1999; Tekle, Tan and Berger, 2008a). To support researchers, statisticians strive to find designs which are efficient, reliable, valid and cost-effective even when the researcher unpurposely misspecifies the model and/or its assumptions due to the complexities of the problem. The search for such "robust" designs is, however, not a trivial one.

This dissertation is about the robustness of design choice relative to incorrect model specifications and assumptions. The following sections present the scope of the research and motivate the choice of this topic by means of an example.

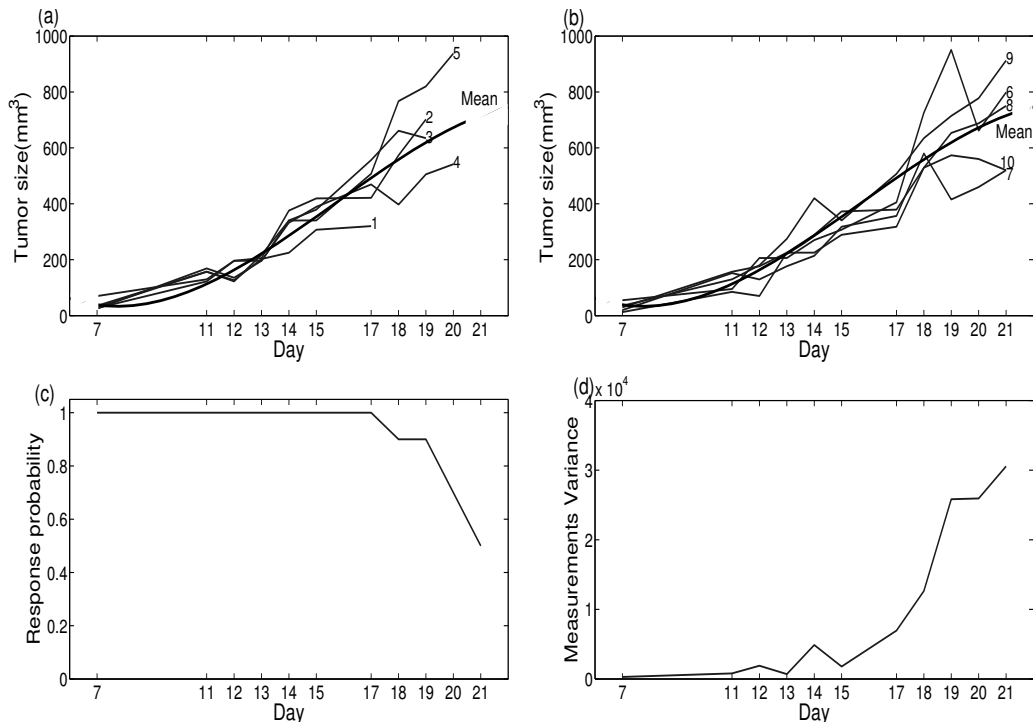
1.1. An example of longitudinal study

The rate of growth of solid tumors, i.e. the time course of metastatic processes is a typical example of a clinical longitudinal study. Oncologists use this variable for diagnosis and prognosis purposes. Tumors are induced in laboratory animals so that the efficacy of potential therapies for malignant diseases can be evaluated. Fukushima et al. (1982a,b) and Tsai and Koziol (1988) reported an experiment conducted in the Cancer Center of the University of California. In this experiment a homogeneous population of mice is injected with mouse colon carcinoma cells. Five days later the presence of the induced tumors was identified and the mice were randomly divided into subgroups. Each group received a different immunotherapy. The sizes of the tumors were observed systematically at 11 time points. The first observation is at day 7 while the last is taken at day 21.

The data of the mice group undergoing treatment an injection of immune RNA is used to illustrate some of the features that characterize longitudinal

data. In Figures 1.1a and 1.1b, the profiles of 10 different mice are shown along with their mean profile. Every mouse profile is denoted by a number from 1 to 10. For the sake of clarity, mice 1 to 5 which drop out before the end of the trial are shown in Figure 1.1a, whereas mice 6 to 10 which complete the whole trial are presented in Figure 1.1b. Both figures show that the growth of the tumor size in the mice treated with immune RNA changes over time and that at the start of the trial mice have about the same tumor size. However, the way the tumor size develops in time varies among mice. Hence, it will be desirable to have a design (and model) which allow a flexible formulation of the overall (mean) population and the individual (subject-specific) behavior. This means that the behavior of the overall mice group as well as the behavior of each mouse separately should be allowed to fit polynomial models of different degrees.

Figure 1.1: Changes over time of the (a) tumor sizes (mm^3) of mice that drop out, (b) tumor sizes (mm^3) of mice that complete the trial, (c) dropout in terms of a response probability function, and (d) estimated total (subject+error) variance(s) of a group of mice receiving injections of tissue culture medium around the growing tumor (immune RNA).



Clinical trials dealing with induced tumors are invasive surveys which

are sensitive to dropout (or attrition). A dropout process is complementary to a response probability, i.e. a high dropout is related to a low response probability and vice versa. In Figure 1.1c, the change of the response probability of the group of mice over time is depicted. The response probability is a monotonic decreasing function with the lowest (marginal) response probability occurring on the last day of the trial. Often researchers will expect that data will be (partially) missing; however, neither the form of the response probability nor the magnitude of the percentage of missing data are known beforehand. As it is shown in Figure 1.1c only half the mice reaches the end of the trial and a quadratic function may be suitable to describe the response probability of the mice. Therefore, it will be convenient to have a design (and model) capable of dealing with uncertainties on the profile and the magnitude of the dropout.

Another feature of longitudinal data concerns the high variability of subjects responses. In Figure 1.1d the change of the measurements variability over time is illustrated. This change of variability of the data has two sources of variation: the between- and the within-subjects variation. The first source of variation reflects the difference between-subjects. That is individual and population mean tumor sizes differ over time. In Figures 1.1a and 1.1b, it can be seen that at the beginning (day 7) the tumor size of every mouse is more or less similar to that of any other mice. Hence, at the beginning of the trial, there is a small between-subjects variance. However, if a model would be fitted to describe the tumor size growth of every mouse separately over time then, the fitted lines will have different slopes. Most of the mice will have high or steep slopes when the tumor grows fast while some of them will have low or unsteep slopes when the tumor remains relatively small. If the tumor sizes vary increasingly over time, then the between-subjects slope variance will increase over time.

The second source of variation corresponds to the error variance within-subjects. The errors at different time points are usually (serially) correlated and the variances of the measurement error increase over time. The former means that there is a correlation between errors over successive time points. The latter means that error variance heteroscedasticity arises over time. In the case of mice tumor growth, Figure 1.1d shows that the change of the total variation of the between- and the within-subject variation may increase exponentially. However, it is often difficult to discriminate which part corresponds to the between- and which to the within-subjects variation. Consequently, a design (and model) which allows the representation of either of these two sorts of variations or a combination of them is desirable.

This example represents the following questions which are interesting

from a design point of view:

1. *How many* time points should be measured and *where* should these points be located to describe the *overall mean and the subject-specific behavior* of the mice group with the *least effort*?
2. If the underlying *model of the tumor growth is incorrectly specified* at the planning stage of the study, does the chosen *design* still provide *enough information* for efficient estimation of model parameters afterwards?
3. *How sensitive is the chosen design to uncertainty about the*
 - *degree of the polynomial model?*
 - *between- and within-subject variation?*
 - *mice dropout?*
 - *change of the within-subject variation (error variance heteroscedasticity) over time?*

From a statistical point of view, a more general question arises:

4. *Are there designs which remain highly efficient in spite of potential departures from the model and its assumptions?*

It is evident from the foregoing questions that a researcher intuitively may question the design choice and may either state or present (explicitly) or expect (implicitly) what properties the design and model must possess. Before dealing with the aforementioned statistical question, background theory on designs, linear mixed models, response probabilities, and variance-covariance matrix estimators is needed.

1.2. Theoretical background

1.2.1. The design

A design to collect longitudinal data is basically the location and allocation of resources: time and subjects. Following Kiefer (1959); Kiefer and Wolfowitz (1960); Kiefer and Wolfowitz (1961); Kiefer (1974), an exact design ξ_N will be defined as follows:

$$\xi_N = \left\{ \begin{array}{cccccc} x_1 & x_2 & \dots & x_j & \dots & x_q \\ n_1 & n_2 & \dots & n_j & \dots & n_q \end{array} \right\}, \quad (1.1)$$

where the integers n_j indicate the number of subjects responding at design points x_j . The subindex j runs from 1 through q , q being the number of design points. The total number of observations is $N = \sum_{j=1}^q n_j$. When data are complete, the number of subjects n_j responding to each x_j is equal, i.e., $n_1 = n_2 = \dots = n_j = \dots = n_q$ and $N = qn_1$. In the event that dropout occurs, the number of subjects responding at x_j will decrease monotonically, i.e., $n_1 > n_2 \geq \dots \geq n_j \geq \dots > n_q$ and the probability of obtaining data at design point x_j will depend upon time.

For the example data in Figure 1.1, the exact design can be written as follows:

$$\xi_N = \left\{ \begin{array}{cccccccccccc} 7 & 11 & 12 & 13 & 14 & 15 & 17 & 18 & 19 & 20 & 21 \\ 10 & 10 & 10 & 10 & 10 & 10 & 10 & 9 & 9 & 7 & 5 \end{array} \right\}. \quad (1.2)$$

Three specific features can be noted in this design with dropout. First, dropout arises at day 18. Second, only 5 mice reach the end of the trial, i.e. day 21. Lastly, the total number of observations is exactly $N = 100$.

1.2.2. The linear mixed model and its assumptions

In general, a model is a parsimonious simplification of reality. It is meant to describe the relation between the response and the most relevant independent variables of a research study. The task of obtaining as parsimonious and meaningful a model as possible, is not an easy one.

The most flexible manner to model longitudinal data is using linear mixed models (Laird and Ware, 1982; Verbeke and Molenberghs, 2000). Let \mathbf{y}_i be a continuous response of subjects $i = 1, \dots, n_1$ with three components: the overall-population effects, the random subject-specific effects and the first-level within-subjects error terms. For each subject, the measured response is given by:

$$\mathbf{y}_i = X_i\boldsymbol{\beta} + Z_i\mathbf{b}_i + \mathbf{e}_i, \quad (1.3)$$

where $\boldsymbol{\beta}$ denotes a $p \times 1$ vector of unknown population parameters, X_i is a known $q \times p$ design matrix linking $\boldsymbol{\beta}$ to \mathbf{y}_i . The number of the fixed parameters of the polynomial regression and the number of design points (or time points) are given by p and q , respectively. Let \mathbf{b}_i denote a $k \times 1$ vector of unknown subject-specific effects and Z_i be a known $q \times k$ design matrix linking \mathbf{b}_i to \mathbf{y}_i . The random-effects parameter vector \mathbf{b}_i represents how the i^{th} subject deviates from the average population and it is normally distributed with mean $\mathbf{0}$ and $k \times k$ covariance matrix D , i.e., $N(\mathbf{0}, D)$. In practice, the most frequently encountered mixed models have subject-specific random in-

tercept and random slope terms (Hedeker et.al, 1989; Hedeker and Gibbons, 1996). Therefore, throughout this thesis, the quadratic and cubic random effects are considered as fixed.

In this dissertation, it is assumed that D is a 2×2 matrix with elements: d_{11} which represents the variance of the random intercepts; d_{12} which refers to the covariance between the random intercepts and random slopes and d_{22} which corresponds to the variance of the random slopes. Notice that the random variables \mathbf{b}_i and \mathbf{e}_i are independent, identically and normally distributed.

The random error of each subject is represented by \mathbf{e}_i . This error is assumed to be normally distributed with mean $\mathbf{0}$ and covariance $\Sigma_i^{1/2} \Psi_i \Sigma_i^{1/2}$. Σ_i is a diagonal matrix with variances $(\sigma_1^2, \dots, \sigma_q^2)$ on the main diagonal. We allow serial correlations between observations at different time-points. This dependence is modeled through the within-subject correlation matrix Ψ_i , i.e. a positive-definite autocorrelation matrix depending on the parameter ρ . The parameter ρ denotes the autocorrelation between the j^{th} and j'^{th} time points x_j and $x_{j'}$, respectively. The elements of the autocorrelation matrix are given by $\psi_{jj'} = \rho^{\text{abs}(x_j - x_{j'})}$, where $j, j' = 1, 2, \dots, q$. Notice that the time points x_j are not necessarily equally spaced. Two cases are distinguished. First, errors \mathbf{e}_i are homogeneous among different time points x_j and $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_q^2 = \sigma^2$. So, $\Sigma_i^{1/2} \Psi_i \Sigma_i^{1/2}$ simplifies to $\sigma^2 \Psi_i$. Second, the error variance function is heterogeneous, i.e. $\Sigma_i^{1/2} \Psi_i \Sigma_i^{1/2}$; where at least two of the diagonal elements of Σ_i are unequal in value. For a list of other possible variance-covariance structures, see (Verbeke and Molenberghs, 2000, p.99-100).

The estimation of the error variance function depends on the specified model. In the case of the tumor growth example displayed in Figure 1.1d of section 1.1, the response variance would be equal to the error variance only if, the specified model is of the fixed-effects kind i.e. $D = 0$. In this case, the (rounded) estimated standard deviations of the errors would be:

$$\hat{\sigma}_j = [17 \quad 28 \quad 44 \quad 26 \quad 70 \quad 42 \quad 83 \quad 112 \quad 161 \quad 161 \quad 175]. \quad (1.4)$$

1.2.3. Probability of response

Let us rewrite equation (1.1):

$$\xi_N = \left\{ \begin{array}{cccccc} x_1 & x_2 & \dots & x_j & \dots & x_q \\ 1 & \frac{n_2}{n_1} & \dots & \frac{n_j}{n_1} & \dots & \frac{n_q}{n_1} \end{array} \right\}. \quad (1.5)$$

At the first design point x_1 of the study, the probability of response is $p(x_1) = 1$. At the j^{th} design point with response probability $p(x_j)$, the expected number of subjects n_j is $n_1 p(x_j)$. Finally, at the last design point x_q , the number of subjects n_q represents the number of subjects completing the experiment. For $j = q$, n_q is the number of subjects responding at all q design points. Assuming that at least one subject is observed at this last point, $n_q \geq 1$, the response probability $p(x_j)$ is then a monotonically decreasing function with $p(x_1) \geq p(x_2) \geq \dots \geq p(x_q)$.

Notice that a response probability $p(x_j)$ at a design point x_j is complementary to a dropout process $1 - p(x_j)$ at the same design point x_j and that an important assumption is that the dropout occurs through a noninformative mechanism, i.e., responses are missing at random (MAR), see e.g. Little and Rubin (1987).

The response probability $p(x_j)$, where $j = 1, \dots, 11$, of the example on tumor growth is displayed in Figure 1.1(c). The (marginal) response probability of the group of mice at day 21 is 0.5, i.e. $p(x_{11}) = 0.5$. Hence the probability that mice responses are measured at the end of the trial is one out of two.

1.2.4. Asymptotic variance-covariance of estimators

Let us introduce the super-index $[j]$ to group subjects having j measurements. Hence, design matrices of the the fixed- and random-effects, i.e. $X^{[j]}$ and $Z^{[j]}$, respectively are connected to the response probability $p(x_j)$.

If the number of subjects having all responses to all q design points is denoted by $m_q = n_q$ and the number of subjects with only $j < q$ responses is denoted by the difference $m_j = n_j - n_{j+1}$, then the number of subjects m_j having j responses is given as:

$$m_j = \begin{cases} n_q & \text{if } j = q, \\ n_j - n_{j+1} & \text{if } j < q. \end{cases} \quad (1.6)$$

The design matrices $X^{[j]}$ and $Z^{[j]}$ have size $j \times p$ and $j \times k$, respectively. The linear mixed model (1.3) for data with dropout and heterogeneous variance error implies a marginal model with mean $X_i \beta$ and covariance matrix:

$$V^{[j]} = Z^{[j]} D Z^{[j]} + \Sigma^{[j]1/2} \Psi^{[j]} \Sigma^{[j]1/2}. \quad (1.7)$$

Notice that the last term of (1.7) reduces to $\sigma^2 \Psi^{[j]}$ when the error variances are homogeneous.

The matrix $V^{[j]}$ is linked to the asymptotic covariance matrix of the best linear unbiased estimator $\hat{\beta}$ and since m_j subjects share identical matrices $X^{[j]}$ and $Z^{[j]}$, the asymptotic covariance matrix of the best linear unbiased estimator $\hat{\beta}$ is given by:

$$\text{Var}(\hat{\beta}) = \left(\sum_{j=1}^q m_j X^{[j]} V^{[j]-1} X^{[j]} \right)^{-1}, \quad (1.8)$$

where m_j follows (1.6) when data encounter dropout and $m_j = n_q$ for all j , when data are complete.

Asymptotically, the inverse of the variance-covariance matrix (1.8) approaches the value of the Fisher information matrix. Small valued asymptotic variance-covariance matrices of the parameter estimators correspond to large valued information matrices. The Fisher information matrix represents the amount of information contained in a specific design to estimate the model parameters β . Good estimators with high information and high precision have small valued (asymptotic) variances.

Optimal designs follow this idea and are generated based on a particular optimality criterion. However, an important feature of optimal designs is that they are optimal only for a specified statistical model M .

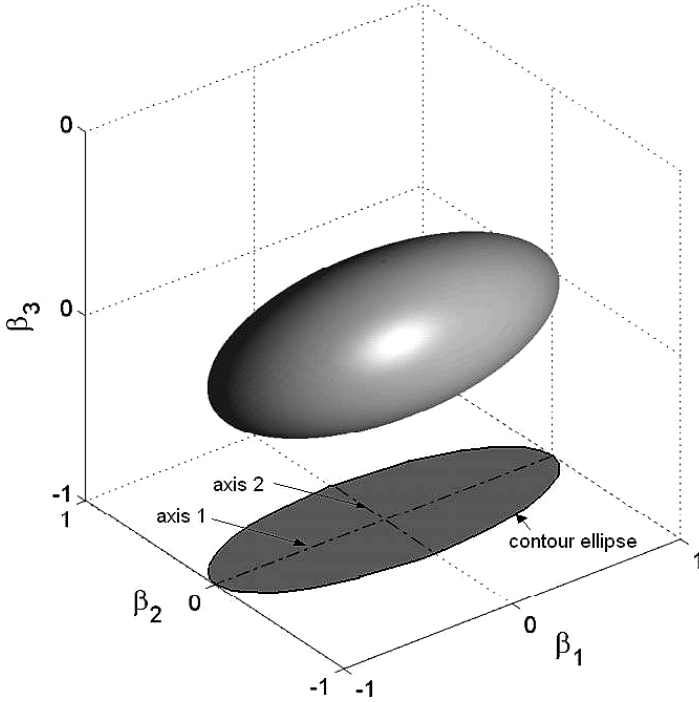
1.2.5. Optimality criteria

There are several optimality criteria to select an optimal design. The most popular criterion is the *D-optimality* criterion; it seeks the *continuous* design that minimizes the determinant of the variance-covariance matrix of the model parameter estimators $\hat{\beta}$, i.e.,

$$\xi^* = \min_{\xi \in \Xi} \left[\text{Det}\{\text{Var}(\hat{\beta})\} \right]. \quad (1.9)$$

The *D-optimal* design ξ^* is chosen among all candidate designs ξ that belong to the set of designs Ξ . ξ^* minimizes the volume of the asymptotic confidence ellipsoid of the parameter estimators. See Figure 1.2, where the simultaneous confidence intervals for two and three parameters, i.e. for $\beta = [\beta_1, \beta_2]^T$ and $\beta = [\beta_1, \beta_2, \beta_3]^T$ are shown. The smaller the volume of this confidence ellipsoid is, the better the estimators of the parameters are. This is an intuitive manner to illustrate the connection between the optimal design ξ^* and the variance-covariance $\text{Var}(\hat{\beta})$. Another advantage of the *D-optimality* criterion is its scale invariance. If a design ξ^* which

Figure 1.2: Simultaneous confidence ellipse and ellipsoid for two and three parameters, where $\beta = [\beta_1, \beta_2]^T$ and $\beta = [\beta_1, \beta_2, \beta_3]^T$, respectively.



is optimal in the time interval $[a, b]$ is needed in another time interval, the linear transformation required to transform the interval $[a, b]$ into $[c, d]$ can be used to transform the design ξ^* into the new time interval (see Ouwers et.al. (2002) for a proof). One of the drawbacks of the D -criterion is related to the fact that the simultaneous confidence interval of the parameters can elongate or stretch out in one specific direction, e.g., the parameter β_2 (axis 2) will be estimated more accurately than the parameter β_1 (axis 1) as can be observed in Figure 1.2. Another disadvantage has to do with the number of parameters in the model. As the complexity of the model increases, so does the number of parameters and finding D -optimal designs demands higher computational effort.

Another criterion is the A -optimality criterion; it seeks to minimize the trace of the variance-covariance matrix of the model parameter estimators, i.e.,

$$\min_{\xi \in \Xi} \left[\text{Trace}\{\text{Var}(\hat{\beta})\} \right]. \quad (1.10)$$

The A -optimal design minimizes the average asymptotic variance of the parameter estimators, i.e. $\min [\text{Var}(\hat{\beta}_1) + \text{Var}(\hat{\beta}_2)]$ for a model with two parameters. The A -optimal design can also be seen as the design that minimizes the sum of the squared lengths of the axes of the ellipse or ellipsoid. Depending on the number of parameter estimators, this sum is proportional to the contour of the ellipse or ellipsoid, as shown in Figure 1.2.

A -optimal designs are characterized because their time points x_j , generally, have unequal weights. This feature could be sometimes considered as a disadvantage. Most important, perhaps is that, A -optimal designs do not remain optimal after a linear transformation of the independent variable(s) scale. This feature is particularly important since longitudinal studies are carried out on a large variety of time intervals. But since A -optimal designs are not scale invariant, comparisons among designs and (more important) among results cannot be done by simple linear transformation of time intervals, designs nor results.

This dissertation focuses on the D - and A -optimality criteria. Among other optimality criteria are the E -, G -, I -, V -optimality criteria. For more details on these criteria, the reader is forwarded to e.g. Kiefer (1959); Kiefer and Wolfowitz (1960, 1961); Fedorov (1972); Kiefer (1974); Pukelsheim (1993); Atkinson and Donev (1996); Dette (1997).

Notice that in this dissertation the optimal designs are discretizations of continuous values; the grid used is quite small and is connected to the non-linear numerical scheme used to find optimal designs.

1.2.6. A manner to investigate robustness of designs

A maximin procedure is used to overcome the problem of local optimality of the encountered optimal designs. The idea is to select a region of interest for each parameter, such that the encountered maximin design will not depend on the individual parameter values. In this manner robust designs that are resistant to deviations of the model assumptions are obtained (Dette and Neugebauer, 1997; Bischoff, 1996).

Robust designs show high relative efficiencies in spite of deviations from their model assumptions. The relative efficiency is a measure to compare designs and it is defined as follows:

$$RE(\xi|\xi^*) = \left[\frac{\text{Det}\{\text{Var}_M(\hat{\beta}|\xi^*)\}}{\text{Det}\{\text{Var}_M(\hat{\beta}|\xi)\}} \right]^{\frac{1}{p}}, \quad (1.11)$$

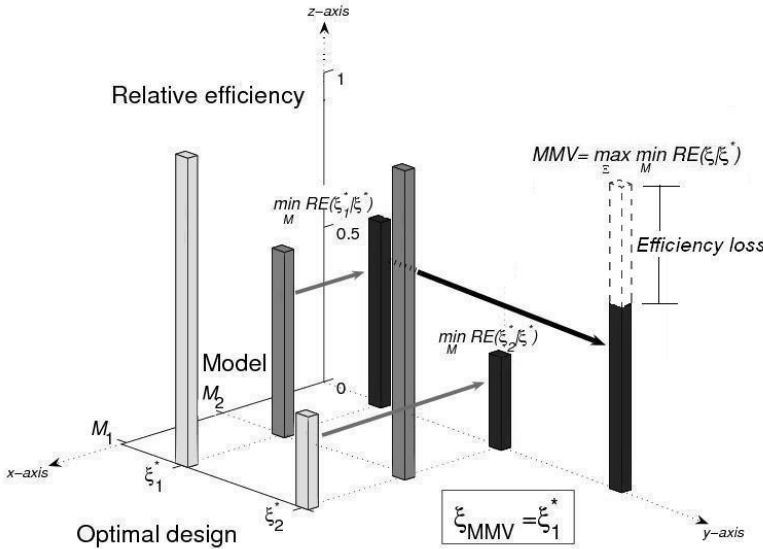
where ξ is a (suboptimal) design and ξ^* is the optimal design, both under

model M . Since the generalized variance of the suboptimal ξ (denominator) will always be larger than that of the optimal design ξ^* (numerator), the values of the relative efficiency are always in the interval $0 \leq RE \leq 1$. The relative efficiency for the D -optimality criterion is raised to the power $1/p$ to and is then independent of the number of fixed parameters p . Moreover, notice that RE s can be interpreted in terms of relative sample sizes. See Atkinson and Donev (1996).

Consider the following example to illustrate how model uncertainty can be overcome. Suppose that a researcher is investigating two particular models to describe tumor growth in mice. One of these two models is correct, but there is no certainty as to which one is. The researcher has obtained two optimal designs using equation (1.9). Each design corresponds to a different model.

Let the x -axis of Figure 1.3 represent the models. Two of these models (connected to their optimal designs) are denoted by M_1 and M_2 . The y -axis represents the "selectable" optimal designs. There are two of those designs, ξ_1^* and ξ_2^* . Since optimal designs are locally optimal, any optimal design will become suboptimal whenever its model assumptions are misspecified. Thus, by using equation (1.11), two trivial and two non-trivial relative efficiencies

Figure 1.3: Illustration of the concept behind the use of a maximin criterion to find robust designs



can be computed. The value of $RE(\xi|\xi^*)$ will be exactly equal to one when, a suboptimal design coincides with the optimal design, i.e. $\xi = \xi^* = \xi_1^*$ or $\xi = \xi^* = \xi_2^*$, otherwise RE values will be smaller than one.

In Figure 1.3, the z -axis gives the relative efficiency while the four computed relative efficiency values are represented by the height of the light-gray and gray bars. The minimum relative efficiency value over the set of "true" models ($M = \{M_1, M_2\}$) is obtained for each "selectable" optimal design. A thick gray-arrow points to these minima, i.e. $\min_M RE(\xi_1^*|\xi^*)$ and $\min_M RE(\xi_2^*|\xi^*)$, respectively. Note that these RE values are displayed (for schematic reasons) as dark-colored bars in Figure 1.3, too. Next, the maximum relative efficiency is selected over all the minimum relative efficiencies of the "selectable" optimal designs, i.e. $\max_{\Xi^*} \min_M RE(\xi|\xi^*)$ where $\Xi^* = \{\xi_1^*, \xi_2^*\}$. This relative efficiency value is called the maximin value (MMV). A thick black-arrow points towards the MMV . In our example, the design $\xi_{MMV}(= \xi_1^*)$ is connected to the MMV and can be called a *robust design* when the MMV has (as a rule of thumb) a value larger than 0.85. This robust design maximizes the minimum relative efficiency due to uncertainty in the underlying model. Consequently, the maximin criterion facilitates a selection of the design whose performance is the best (over the least efficient designs) given a set of candidate models.

In general, the sample size is related to the relative efficiency: $N_{extra} = N(1/RE - 1) \times 100\%$, where N_{extra} is the number of extra required observations such that a (suboptimal)design performs as well as another (optimal/robust)design.

1.3. Robustness of designs

The term robust is used extensively in literature regarding statistics. The following examples are found: robust estimates, robust tests, robust regression, robust procedures, robust algorithms, robust correlation, robust covariance and robust designs. Huber (2004) defines the associated concept of robustness as the 'insensitivity to small deviations from the assumptions'.

In this dissertation, Huber's definition will be adopted and used as the property that represents 'strength' or 'insensitivity' of a design with respect to (unforeseen) departures from the *fixed-effects model and from the assumed values of variance components, and incomplete data due to dropout*. A design will be called *robust* if it is capable of coping well with departures from its model, variance and completeness assumptions in its operating environment i.e., a robust design leads to a minimal loss of efficiency and thus to a minimal

change of its sample size under model misspecifications.

In this dissertation we will focus on the robustness of D -optimal designs against the incorrect specification of (1) the degree of the polynomial function, (2) the variance-covariance matrix of the random effects, (3) the error variance structure, (4) the dropout, and the robustness of A -optimal designs against incorrect assumptions on weights and design scales.

1.3.1. Robustness of D -optimal designs against the incorrect specification of the degree of the polynomial model

During the planning process of a study, researchers need to adopt a model which hopefully gives a good description of the *true underlying model*, i.e. it describes the process best. One cannot assume that it is exactly correct. It can only be assumed that a chosen design provides a reasonably high efficiency for the *assumed model*.

From previous studies (Tan and Berger, 1999; Tekle, Tan and Berger, 2008a), it is known that the specification of the degree of the polynomial model, i.e. the number of fixed parameters is crucial for the identification of a highly efficient design. They suggest choosing a number of design points, as close as possible, to the number of fixed parameters. For fixed-effects and random intercept models, the rule of thumb is that the number of design points should be, ideally, equal to the number of fixed parameters plus one.

Thus, a major concern of a researcher will be *whether the degree of the polynomial of a linear mixed model is incorrectly specified*. In the example in section 1.1, the tumor size increase over time can be specified as linear but it may also be quadratic or even cubic. A chosen design must be able to deal with such a model uncertainty.

1.3.2. Robustness of D -optimal designs against the incorrect specification of the variance-covariance matrix of the random effects

It is known that every subject changes over time in a different manner. Some subjects will show linear-like behavior while other subjects will show curve-like behavior. The main issue is then how to select a design when the assumptions made on the variance-covariance matrix of the random parameters are uncertain.

There are several possible scenarios. One option is to neglect random variations and thus random variances in the model. The model then reduces to the well-known fixed-effects model. A second option is to assume that subjects differ only at start-up. The variance between the subjects'

intercepts is represented by the random intercept variance and this structure is known as the random-intercept model. Another option is to assume that subjects will differ at the beginning of the study and their behavior will also change over time (i.e., by having a different slope). The variation among subjects slope will be denoted by a random slope variance. This is the random intercept- and random-slope model. In this case, it is implicitly assumed that the random intercept variance and the random slope variance are uncorrelated. However, this may not be always the case and correlations between random intercept and random slope variances are then taken into account in the model.

A major issue will then be *whether the between-subjects variance structure is misspecified*. In the tumor growth (see Figure 1.1a and 1.1b) there is a large variation among mice over time. At day 7 all mice have about the same intercept value, i.e. the between-subjects variance is small. However, the mice individual slopes display a large spread at day 21, i.e. the between-subjects—variance over time may become large. So, choosing a design optimal for a fixed-effects model rather than a random-intercept and random-slope model may affect the accuracy of the estimated model parameters.

1.3.3. Robustness of D -optimal designs against the incorrect specification of the error variance-covariance structure

Another issue to be addressed when selecting a design is related to the assumptions regarding the error variance parameters and the error variance structure.

The simplest but most unlikely structure is the uncorrelated one with constant error variance, where it is assumed that the measurements of the subjects are all independent over time and the error variances are constant. A second structure corresponds to an exponentially decreasing function known as an autoregressive correlation structure. It means that there is a time-lag between time points and the error variance remains constant. There is also the heterogeneous autoregressive correlation structure where the error variance is heterogeneous, i.e. the error variances change over time while the serial correlation decreases over time.

In this case, the major concern of a researcher will be *what-if the within-subject variance-covariance structure is misspecified*. In the example, see Figure 1.1d, it is shown that the variability of the responses at design points changes over time. Part of this variability can be explained by the heterogeneous within-subject variance.

1.3.4. Robustness of D –optimal designs against the dropout

When carrying out a longitudinal study, subjects might drop-out of the study at particular time points. Subjects will be missing due to different circumstances, often beyond the control of the researcher. Further efforts on keeping these subjects in the study may become very expensive or simply impossible.

To analyse data with dropout it is common to use the direct (maximum likelihood) estimation of the inverse of the information matrix. Likelihood-based methods are broadly valid, since they only require that the missing data mechanism to be missing at random (MAR) rather than missing completely at random (MCAR) (Molenberghs and Kenward, 2007).

A missing completely at random (MCAR) process assumes that the probability of observing a response does not depend neither on the observed data nor on the unobserved data. Moreover, if these probabilities depend only on the observed data, then the non-response pattern is said to be missing at random (MAR). These two types of non-response are considered as *ignorable or non-informative*. The maximum likelihood estimates for the parameters of the distribution of the response do not depend on the model for the non-reponse pattern and therefore can be ignored. The obtained estimators will be unbiased but less efficient. For further details on missingness patterns see e.g. Little and Rubin (1987); Diggle et al. (1994), Everitt (1998); Molenberghs and Kenward (2007).

The major concern of a researcher will be *whether there is dropout and if the design has taken this into account*. In the example, the number of repeated measurements of the tumor volume vary across the mice. See Figures 1.1a and 1.1c. Notice that researchers usually will not be aware of the missing pattern mechanism beforehand. Hence, any chosen design must be able to deal satisfactorily with uncertainty regarding the probability of response.

1.3.5. Robustness of A –optimal designs against incorrect assumptions on weights and design scales

Up to now, the researcher’s goal included the use of the D –optimality criterion, i.e., a design with a minimum generalized variance of the parameter estimators $\hat{\beta}$.

The researcher’s goal may be to target the minimum average of the variance of the parameter estimators rather than that of the determinant. The A –criterion is, perhaps, more appealing to practitioners because it is based on the idea of minimizing the average variance of the parameter

estimators of the model. Moreover, A -optimal designs are, in contrast to D -optimal designs, more often characterized by having unequal weights, i.e. they are unbalanced designs.

Two issues that may concern the researcher are *whether A -optimal designs with the constraint of equal weights are used rather than their unequal-weight counterparts?* and *whether A -optimal designs obtained for the time interval $[a, b]$ are used in the interval $[c, d]$?, (where $a \neq c$ and/or $b \neq d$).*

It must be pointed out that, the reader might encounter terms like *distributional robustness* and *noise robustness* in statistical literature. The term distributionally robust (or outlier resistant) has been coined to describe procedures that safeguard estimates from distributional deviations. There are several diagnostics and procedures to deal with this problem. For further details, see Tukey (1960); Huber (2004). The term noise robustness refers to a procedure that allows to design a product, in such a way, that its performance (as measured by a quality characteristic) is *insensitive* to variation in the noise factors (external variables) (Huele, 1998). The so-called Taguchi designs are connected to this type of robustness. This dissertation is not about these types of robustness because they have been treated extensively elsewhere. See for example Pearson (1895); Snedecor (1937), Scheffé (1959) and Taguchi (1987); Moen et. al. (1991); Montgomery (2000), Rao et. al. (2008), respectively.

1.4. The goal

This dissertation will focus on *finding robust designs for longitudinal studies* in which the actual measurement of individuals takes place at several *well defined* occasions and that these *well defined* occasions are selected following a criterion that best suits the researcher's goal and *ideally minimize the effect of any violations of the model and its assumptions*.

1.5. Organization of the thesis

This dissertation contains six self-contained chapters, a chapter containing overall conclusions and a summary. Notation and formulae have been presented as much as possible as is standard in statistical practice; when necessary additional symbols or equations have been added.

Chapter 2 is an extension to the work of Tan and Berger (1999); Ouwens et.al. (2002). It contains a systematic approach to find highly efficient designs when the model and the variance-covariance of the model are incorrectly specified. A list of highly robust designs is provided.

Chapter 3 tackles the issue of increasing error variances over time. So far, this topic has been studied for the case of independent errors only. In this chapter, optimal designs for linear mixed models with heterogeneous error structure are provided.

Chapter 4 considers the topic of dropout, i.e. subjects dropping out at random. The chapter focuses on finding optimal designs for linear mixed models when dropout arises but has not been accounted for in the design stage. Moreover, designs for two sorts of response probability functions are discussed.

Chapter 5 is devoted to the investigation of two features of the A -optimality criterion. In this chapter, the consequences of using the A -optimal design with incorrectly assumed equal weights are shown. Thereafter, the loss of efficiency of assuming incorrectly that A - designs can be scaled into four different design intervals by using a linear transformation is studied.

Chapter 6 introduces the tool ODMIXED which enables the user to find optimal designs when the model assumptions are misspecified. Two examples connected to chapters 3 and 4 have been worked out to explain how to obtain optimal designs in two specific situations and how to use this software to compare alternative designs. A list of generated files is included to ease the manipulation of outcomes.

Chapter 7 lists the conclusions of this dissertation and suggests potential spin-offs of this research. The summary states the motivation about seeking robust designs, the obtained results and our conclusions in a condensed form.

CHAPTER 2

Highly efficient designs to handle the incorrect specification
of linear mixed models ¹

¹ *Communications in Statistics - Simulation and Computation*(2009)**38**:1,14–30.

Abstract

We apply a maximin criterion to examine the relative efficiency of several D_q -optimal designs for a family of linear mixed models. Incorrect specifications of the order of the polynomial, size of the autocorrelation parameter, number of random parameters, and the correlation between random intercept and random slope are investigated. The results of our study allow us to draw the following conclusions: 1) the maximin D_q -optimal design encountered appears to be highly efficient; 2) the variation of the minimum relative efficiencies of D_q -optimal designs of the family of linear mixed models that were studied, decreases as the order of the polynomial increases; 3) the effect of the autocorrelation parameter on the relative efficiencies of D_q -optimal designs is the largest for first-degree polynomials; and 4) the relative efficiency of the equidistant design is lower than the reference value 0.85.

Keywords: D -optimality; D_q -optimality; first-order autocorrelation; linear mixed models; maximin criterion; relative efficiency; robust designs

2.1. Introduction

Longitudinal and repeated measure designs are widely used by practitioners for experimental and observational studies. These designs target to investigate changes over time in subjects and are usually analyzed by linear mixed models (LMM). See e.g. Laird and Ware (1982) and Verbeke and Molenberghs (2000).

An optimal design for longitudinal and repeated measurements depends on the assumed model. Practitioners, however, do not know beforehand the underlying model of a studied phenomenon and data analysts usually choose designs with sufficient number of time points at convenient places. This often results in loss of efficiency. The estimated coefficients of the underlying model are often not estimated efficiently because a chosen model with covariance matrix may differ from the true model and covariance matrix. In this paper, we will focus on the problem of finding efficient designs for linear mixed models that remain highly efficient even if the true model is not known.

The importance of studying optimal designs for linear mixed models has been advocated by many authors (Abt et.al., 1997, 1998). These authors searched for optimal designs for linear and quadratic growth models with random intercept and autocorrelated structures.

The robustness of designs against incorrect specifications of the polynomial model has been studied in recent years by Wong (1994); Dette (1997) and Moerbeek (2004). Few authors, however, have provided evidence of the drawbacks of specifying incorrectly random-effects models. Tan and Berger (1999); Berger et.al (2002) and Berger and Tan (2004) compared some D -optimal designs for polynomial regression models where either the random intercept variance or the random slope variance was fixed given a first-order autocorrelation structure, $AR(1)$. They showed that regardless the underlying polynomial regression model, the number of repeated measures should be chosen as close as possible to the number of regression parameters.

Berger et.al (2002), Ouwens et.al. (2002) and Berger and Tan (2004) studied maximin designs for mixed models. Their papers focuss on D -optimal cohort designs for first- and second-degree polynomial models with a (fixed) value for either the random intercept or the random slope variance and within specific intervals for the serial correlation parameter and the error variance.

In this paper we will expand their results and search systematically for efficient designs for a larger range of linear mixed models. The issues of concern will be 1. expand the class of models to a third-degree polynomial

model, 2. extend the parameter values of both: the random intercept and the random slope variances, and 3. show that a discrete approximation of the standardized maximin method proposed by Dette and Neugebauer (1997) and Dette et.al. (2006) will lead to highly efficient designs for models and combinations of parameter values which often occur in practice.

The remainder of this paper is as follows. In section 2.2, we will include a motivating example. In section 2.3, we will introduce the linear mixed model framework and the formulation of the maximin criterion for these models. In section 2.4, we will explain our numerical approach. Thereafter, we will present the obtained results in section 2.5. We will discuss some special cases in section 2.6. Finally, in section 2.7 we will summarize our conclusions.

2.2. Motivating example

In longitudinal studies of psychiatric disorders, uncovering relationships between pharmacological treatments and its therapeutic effects in patients is decisive. An experiment involving an antidepressive drug treatment for depressive inpatients was conducted to examine the relationship between antidepressants plasma levels and severity of depression. See for details Gram et.al. (1976) and Reisby et.al (1977). The inpatients were followed during an one-week baseline placebo period. Thereafter, a fixed drug's dose was administered during four weeks. Plasma-levels of the drug in blood samples were measured and the clinical response of patients was rated using the Hamilton Rating Scales for Depression (HAM-D) at a total of six equally spaced time points.

Notice that this problem deals with time and a time-dependent covariate (plasma-level). This chapter deals only with time, however, similar approach might be extended to studies dealing with time-dependent covariates.

Hedeker et.al (1989) and Hedeker and Gibbons (1996) used linear mixed models to analyze the experimental data and regress HAM-D scores on weeks by polynomial models with random intercepts and random slopes. Their results indicated a significant effect of the treatment on the response across the time points and evidence of a first-order autoregressive process in the residual errors.

The design of this study is characterized by the number of time points, the location of the time points and by the number of patients measured at the time points. In general, the design of such studies can be improved if more specific information is available about the best fitting linear mixed model. Information about the degree of the polynomial describing the trend over time, the random parameters and the autocorrelation errors may enable the

researcher to choose a locally optimal design for such studies. The problem is, however, that such information is rarely available in the design stage of a study. Sometimes, previous research may provide reasonable guesses about these parameter values, but usually the researcher does not know the information required to find an optimal design for his/her study. The problem of finding an optimal design for linear mixed models is hampered by the fact that the information about the parameter may be incorrectly specified. A robust design for these studies should be capable to deal with the incorrect specification of the degree of the polynomial involved, the random variance-covariance structure and the size of the serial correlation. The use of a maximin criterion to identify maximin designs would be an alternative to overcome model and covariance structure uncertainty and reduce experimental costs.

2.3. Linear mixed models and optimality of designs

Let \mathbf{y}_i be a continuous response of subjects $i = 1, \dots, n$ with three components: the overall-population effects, the random subject-specific effects and the first level within-subjects error terms. For each subject, the measured response is given by:

$$\mathbf{y}_i = X_i\boldsymbol{\beta} + Z_i\mathbf{b}_i + \mathbf{e}_i, \quad (2.1)$$

where $\boldsymbol{\beta}$ denotes a $p \times 1$ vector of unknown population parameters, X_i is a known $q \times p$ design matrix linking $\boldsymbol{\beta}$ to \mathbf{y}_i . The number of the fixed parameters of the polynomial regression and the number of design points (or time points) are given by p and q , respectively. Let \mathbf{b}_i denote a $k \times 1$ vector of unknown subject-specific effects and Z_i be a known $q \times k$ design matrix linking \mathbf{b}_i to \mathbf{y}_i . The random-effects parameter vector \mathbf{b}_i represents how the i^{th} subject deviates from the average population and it is normally distributed with mean $\mathbf{0}$ and $k \times k$ covariance matrix D , i.e., $N(\mathbf{0}, D)$. We will restrict ourselves to a 2×2 covariance matrix D , where d_{11} stands for the variance of the random intercepts, d_{12} is the covariance between the random intercepts and random slopes and d_{22} corresponds to the variance of the random slopes. In practice, the most often encountered mixed models have only random intercept and random slope terms (Hedeker et.al., 1989; Hedeker and Gibbons, 1996). Thus, the quadratic and cubic random effects are assumed to be fixed, throughout.

The errors \mathbf{e}_i are distributed as $N(\mathbf{0}, \sigma^2\boldsymbol{\Psi}_i)$. Here $\sigma^2\boldsymbol{\Psi}_i$ is a positive-definite covariance matrix, and depends on the correlation parameter ρ of the time lag between the j^{th} and j'^{th} time points x_j and $x_{j'}$, i.e., $\rho^{\text{abs}(x_j - x_{j'})}$.

These elements of the within-subject correlation matrix Ψ_i correspond to an exponentially decreasing function known as autoregressive correlation structure in which the time points x_j are not necessarily equally spaced. We will symbolize this structure by AR1. Although other structures for Ψ_i exist, we will restrict ourselves to this structure because it is the most often encountered structure in longitudinal designs. For a list of other possible variance-covariance structures, see Verbeke and Molenberghs (2000, p.99-100). Notice that the random variables \mathbf{b}_i and \mathbf{e}_i are independent, identically and normally distributed.

In what follows we will assume that design matrices X_i and Z_i and covariance matrix V_i are the same for each subject. So, if we will consider that measurements are taken at the same design points for all subjects, the model (2.1) can be written as a family of linear mixed models: $\mathbf{Y} = X\beta + Z\mathbf{b} + \mathbf{e}$, whereof \mathbf{Y} , X , β , Z , \mathbf{b} , and \mathbf{e} are all matrices concatenating n subject-specific regression models, i.e. $\mathbf{Y} = (\mathbf{y}'_1, \dots, \mathbf{y}'_n)'$, $X = (X_1, \dots, X_n)'$, $\beta = (\beta_1, \dots, \beta_p)'$, $Z = \text{Diag}(Z_1, \dots, Z_n)$, $\mathbf{b} = (\mathbf{b}'_1, \dots, \mathbf{b}'_n)'$ and $\mathbf{e} = (\mathbf{e}'_1, \dots, \mathbf{e}'_n)'$. This linear mixed model implies a marginal model \mathbf{Y} distributed normally with mean $X\beta$ and covariance matrix:

$$V = ZDZ' + \sigma^2\Psi, \quad (2.2)$$

where $D = \text{Diag}(D, \dots, D)$ and $\Psi = \text{Diag}(\Psi_1, \dots, \Psi_n)$. The matrix product ZDZ' and matrix Ψ have size $(nq \times nq)$.

The best linear unbiased estimator of β has a variance-covariance matrix

$$\text{Var}(\hat{\beta}) = (X'V^{-1}X)^{-1}, \quad (2.3)$$

which is equal to the inverse Fisher information matrix. Small valued asymptotic covariance matrices of the parameter estimates correspond to large valued information matrices. Fisher information matrix represents the amount of information contained in a specific design and expresses the amount of precision by which the model parameters β are estimated by $\hat{\beta}$. Good estimators with high precision have small valued estimated (asymptotic) variances, $\text{Var}(\hat{\beta})$.

2.3.1. Optimal designs

In this section, we introduce a design space Ξ and a parameter space Ω as the building blocks required to handle and apply optimality and maximin criteria.

Design space, Ξ

Let us introduce a design ξ in the design space Ξ . ξ is a probability measure with finite support in x_j , where $j = 1, \dots, q$ and without loss of generality $-1 \leq x_j \leq 1$; that is, ξ represents the relative proportion of the total observations taken at x_j 's time points (Kiefer, 1959). The design ξ is defined as:

$$\xi = \left\{ \begin{array}{cccccc} x_1 & x_2 & \dots & x_j & \dots & x_q \\ 1/q & 1/q & \dots & 1/q & \dots & 1/q \end{array} \right\}, \text{ where } (1 \leq j \leq q). \quad (2.4)$$

where $\int_{-1}^1 \xi(dx) = 1$ and the weights are all equal to $1/q$. In our study we have considered data having equal weights, meaning that no missing patterns arise. Unequal weights may arise in practice and are known as unbalanced designs.

Parameter space Ω

Let us define the parameter space Ω as the parameters connected to the family of marginal linear mixed models in (2.1):

$$\begin{aligned} \Omega = \{ \omega = (q, p, \beta, D, \rho, \sigma^2) : 2 \leq p \leq q, \beta \in \mathcal{R}^p, \\ D \geq \mathbf{0}, 0 \leq \rho < 1; \sigma^2 > 0 \} \end{aligned} \quad (2.5)$$

Ω contains the mean parameters β , the variance and covariance parameters of D and Ψ , and the variance of the errors, σ^2 .

We consider the space Ω to have a total of W different combinations of parameters. Let us denote $w = 1, \dots, W$ to identify each possible combination of parameters in Ω . There is a one-to-one correspondence between each element ω of the parameter space Ω and the model M_w . Each combination of parameter values ω determines a linear mixed model of the form (2.1) and for each model M_w we can derive an optimal design.

Notice that quadratic and cubic-effects are kept as fixed. In practice, one often only specify deviations from the average that change linearly in time. Quadratic- and cubic-effects could also be considered, but its existence and significance might be much harder to show because of lack of power. See Verbeke and Molenberghs (2000, p.70-71).

Optimality criterion

Different optimality criteria are proposed in the literature to compare competing designs (Kiefer, 1959; Kiefer and Wolfowitz, 1961; Fedorov, 1972;

Atkinson and Donev, 1996). In this paper, we will focus on the determinant criterion or D -optimality criterion. The idea is to minimize the determinant function of the asymptotic covariance matrix for the estimated parameter $\hat{\beta}$ for a given model M_w , i.e.,

$$\min_{\xi \in \Xi} \Phi\{\text{Var}_{M_w}(\hat{\beta})\}. \quad (2.6)$$

This criterion is particularly interesting since it has a natural and intuitive interpretation among practitioners as the minimum volume of the confidence ellipsoid about β . Another important advantage is that the D -optimality criterion is invariant with respect to reparametrization (Ouwens et.al., 2002).

The determinant of the $\text{Var}(\hat{\beta})$ can be written as,

$$\begin{aligned} \text{Det}\{\text{Var}_{M_w}(\hat{\beta}|\xi)\} &= \text{Det}\left(X'_\xi V_\xi^{-1} X_\xi\right)^{-1}, \\ &= \text{Det}\{X'_\xi (Z_\xi D Z'_\xi + \sigma^2 \Psi_\xi)^{-1} X_\xi\}^{-1}, \end{aligned} \quad (2.7)$$

where ξ refers to the design that is used for the estimation of $\text{Var}(\hat{\beta})$ under model M_w .

By applying relation (2.6) to (2.7), we obtain that the D -optimal design ξ^* is the design that satisfies:

$$\text{Det}\{\text{Var}_{M_w}(\hat{\beta}|\xi^*)\} \leq \text{Det}\{\text{Var}_{M_w}(\hat{\beta}|\xi)\} \quad \forall \xi \in \Xi, \quad (2.8)$$

where $\hat{\beta}$ is the vector of the estimated parameters for designs ξ^* and ξ under model M_w , respectively.

A complete description about approximate theory on the general equivalence theory for optimum designs is provided in Fedorov (1972); Atkinson and Donev (1996); Kiefer (1974) and Pukelsheim (1993).

2.3.2. Maximin criterion

Different approaches can be used to investigate efficiency loss when a model is incorrectly specified; among them are the sequential approach (Wynn, 1970) and the Bayesian approach (Chaloner and Verdinelli, 1995). For a review of these methods, the reader is referred to Atkinson and Donev (1996). In this paper we will focus on the maximin approach.

The maximin approach was used to overcome the problem of local optimality (Dette and Neugebauer, 1997; Muller, 1995; Bischoff, 1996). In this approach practitioners and analysts select a region of interest for each parameter, such that maximin designs do not depend on the individual pa-

parameter values. Thus, the estimation of the variance random components is not covered in this manuscript but a set of values is considered instead.

The maximin designs in this paper are based on D -optimal designs. The space of all D -optimal designs $\xi_\omega^*(w = 1, \dots, W)$ for all linear mixed models is Ξ^* ; all ξ_ω^* are elements of Ξ^* and Ξ^* is a subset of Ξ , i.e., $\xi_\omega^* \in \Xi^* \subset \Xi$. Two types of D -optimal designs ξ_ω^* can be distinguished:

$$\begin{aligned}\xi_S^* &:= D\text{-optimal design for a selected model } M_S, \\ \xi_T^* &:= D\text{-optimal design for a true model } M_T.\end{aligned}$$

Notice that the D -optimal design space Ξ^* includes the D -optimal designs ξ_S^* and ξ_T^* for the selected model M_S and true model M_T , respectively.

The only parameters that are fixed in models M_S and M_T are the number of points $q = 4$, the residual variance $\sigma^2 = 1$ and the interval $[-1, 1]$. The polynomial degree $(p - 1)$, the variance of the random intercept d_{11} , the variance of the random slope d_{22} , the autocorrelation parameter ρ and the covariance d_{12} between random parameters under the selected model M_S and the true model M_T are not necessarily the same.

The maximin procedure is implemented in three steps. First, we compare the standardized relative efficiencies of an optimal design ξ_S^* for a selected model with respect to the efficiencies of the optimal designs ξ_T^* for a set of assumed true models. Second, we seek for the minimum of all these relative efficiencies over all true (assumed) models, i.e., $M_T = M_w \in \{M_1, \dots, M_W\}$. Finally, the maximum of all the minimum relative efficiencies is selected over all possible D -optimal designs for the selectable models, i.e., $M_S = M_w \in \{M_1, \dots, M_W\}$.

The first step is the most cumbersome and can be illustrated as follows. For each assumed true model M_T , the generalized variance of the estimated regression coefficients $\text{Var}_{M_T}(\hat{\beta}|\xi_T^*)$ given the optimal design for the true model is computed, and compared to the generalized variance of the estimated regression coefficients $\text{Var}_{M_T}(\hat{\beta}|\xi_S^*)$ given the optimal design of a selected model.

The relative efficiency of the selected D -optimal design compared to the true optimal design is:

$$\begin{aligned}
RE(\xi_S^*|\xi_T^*) &= \left[\frac{\text{Det}\{\text{Var}_{M_T}(\hat{\beta}|\xi_T^*)\}}{\text{Det}\{\text{Var}_{M_T}(\hat{\beta}|\xi_S^*)\}} \right]^{\frac{1}{p_T}} \\
&= \left\{ \frac{\text{Det}(X_{\xi_T^*}' \hat{V}_{\xi_T^*}^{-1} X_{\xi_T^*})^{-1}}{\text{Det}(X_{\xi_S^*}' \hat{V}_{\xi_S^*}^{-1} X_{\xi_S^*})^{-1}} \right\}^{\frac{1}{p_T}}, \tag{2.9}
\end{aligned}$$

where $\hat{V}_{\xi_S^*}$ and $\hat{V}_{\xi_T^*}$ are the estimated variance-covariance matrix of the selected optimal design ξ_S^* and true optimal design ξ_T^* both under model M_T . Notice that the D -optimal design ξ_S^* is optimal for the selected model M_S , but not for the true model M_T and that the numerator of (2.9) has a smaller variance than the denominator when the optimal design ξ_S^* for the selected model is not equal to ξ_T^* for the true model. The number of fixed parameters of the true model is equal to p_T . Taking the ratio of the determinant in (2.9) to the $1/p_T$ power results in an efficiency measure which is proportional to design size, irrespective of the dimension of the model (Pukelsheim, 1993; Atkinson and Donev, 1996). The interpretation of the relative efficiency in (2.9) is straightforward. Two replicates of a selected optimal design ξ_S^* for which $RE = 0.5$ would be as efficient as one replicate of the true optimal design ξ_T^* . Relative efficiencies of 0.85 or higher are usually preferred and the relative efficiency of an optimal design with respect to itself is clearly equal to 1. Notice that equation (2.9) can be also used to compare designs with different number of repeated measures (Ouwens et.al., 2002, p.736).

For the last step, the maximin value MMV requires to be defined in terms of relative efficiency, i.e.,

$$MMV = \max_{\xi_S^* \in \Xi^*} [\min_{\xi_T^* \in \Xi^*} RE(\xi_S^*|\xi_T^*)]. \tag{2.10}$$

For each ξ_S^* in the design space Ξ^* , the minimum of relative efficiencies $RE(\xi_S^*|\xi_T^*)$ is selected over all ξ_T^* designs. Then the maximum of all minima is selected over all ξ_S^* designs. The design that has this maximin value MMV is referred to as the maximin D_q -optimal design ξ_{MMV} . Notice that this paper focus on model misspecification, thus the number of design points are the same for the selected and the true model and we will only compare design with the same number of time points.

D_q -optimal designs

' D_q -optimal' design refers to D -optimal designs where the number of time points is fixed at $q \geq p$, the number of regression parameters. A design ξ_q^* is D_q -optimal if $\text{Det}\{\text{Var}_{M_w}(\hat{\beta}|\xi_q^*)\} \leq \text{Det}\{\text{Var}_{M_w}(\hat{\beta}|\xi_q)\}$, for all $\xi_q \in \Xi$. For further details on these optimal designs, the reader is referred to Tan and Berger (1999).

In general, notice that, we will refer to D_q designs as the local optimum design if Model M_w where $w = 1, \dots, W$ is used. Later on, the reader will notice that the subindex w is used to point to a true model M_T with optimal design ξ_T^* or to the selected model M_S with optimal design ξ_S^* .

2.4. Numerical analysis

In this section, we tackle numerically the problem of finding efficient designs that remain efficient even if the model is misspecified.

It is well-known that the correct specification of the linear mixed models depends upon four sources of uncertainty, namely the order of the polynomial ($p - 1$), the size of the autocorrelation coefficient ρ , the number k of random effects, and the correlation ρ^* between the random intercept and random slope.

To deal with uncertainty about the order of the polynomial, the models M_w will be restricted to the set $S_p = \{2, 3, 4\}$, i.e., the first-degree ($p = 2$, linear), the second-degree ($p = 3$, quadratic) and the third-degree ($p = 4$, cubic). These polynomial models are mostly encountered in longitudinal studies in health sciences and medicine (Hedeker et.al., 1989; Hedeker and Gibbons, 1996; Keller et.al., 2003). The number of time points chosen is $q = 4$, it enables us to test the goodness-of-fit. Higher (fourth or fifth) order polynomials are not considered because they may not be always meaningfully interpreted in any biological or physical sense.

To deal with the uncertainty of the size of the autocorrelation coefficient ρ , a convenient grid is used that is $S_\rho = \{0.0001, 0.1, 0.2, 0.3, 0.6, 0.9\}$. Since ρ is chosen from the set S_ρ , six different within-subject residual variance matrices Ψ can be specified.

To handle uncertainty of the random effects, four classes of linear mixed models are considered. These models are based on the number of random effects $k \in S_k = \{0, 1, 2\}$:

1. The case $k = 0$ leads to fixed effect models with AR1 structure, i.e. there is no random term in the variance covariance matrix V . The

random variance-covariance matrix D is $\mathbf{0}$ in block matrix \mathbf{D} of equation (2.2) and the matrix V depends only on matrix Ψ . This class of fixed models will be denoted by **AR1**.

2. The case $k = 1$ enhances random intercept models with AR1 structure. This class of models is denoted by **RI-AR1**. Because, besides the autocorrelated structure for the residuals, there is also a random intercept (RI) variance in matrix $D = (d_{11})$. Random intercept models with AR1 structure are obtained by choosing a value d_{11} from the set of random intercept variances $S_{d_{11}} = \{1, 2, 3, 4, 10\}$ and a value for $\rho \in S_\rho$.
3. The case $k = 2$ allows random intercept and random slope models with AR1 structure. This class of models is denoted by **RI-RS-AR1**. The random intercept (RI) variance and random slope (RS) variance of matrix D are both non-zero, i.e. $D = \text{Diag}(d_{11}, d_{22})$ and residuals are autocorrelated. This model is constructed by choosing a value for the random intercept variance $d_{11} \in S_{d_{11}}$, a value for the random slope variance d_{22} from the set of random slope variances $S_{d_{22}} = \{1, 3\}$ and a value for $\rho \in S_\rho$.
4. In this case, $k = 2$ and the covariance between random intercept and random slope is non-zero. This covariance depends on the correlation ρ^* between the two random variance parameters. The main diagonal elements of $D = \begin{pmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{pmatrix}$ are filled by values $d_{11} \in S_{d_{11}}$ and $d_{22} \in S_{d_{22}}$. Then, using the relation $d_{12} = \rho^* \sqrt{d_{11} d_{22}}$ the off-diagonal elements of the symmetric matrix D are fixed. Notice that D -optimal designs are invariant with respect to the sign of d_{12} (Ouwens et.al., 2002). The correlation ρ^* is chosen from a set $S_{\rho^*} = \{0.5, 0.8\}$ and the parameter ρ is chosen from S_ρ . This class of models is denoted by **RI-RS- ρ^* -AR1**.

In this paper, we have neglected the fact that the parameters of the V matrix need to be estimated. The optimal designs are calculated conditional on the parameters of the V matrix, which is often done in literature (Näther, 1985, p.9).

Given the sets of parameter values $S_p, S_\rho, S_k, S_{d_{11}}, S_{d_{22}}$ and S_{ρ^*} , the parameter space of our numerical analysis is defined as:

$$\begin{aligned} \Omega = \{ \omega = (q, p, \beta, D, \rho, \sigma^2) : q = 4, p \in S_p, \beta \in \mathcal{R}^p, \\ d_{11} \in S_{d_{11}}, d_{22} \in S_{d_{22}}; \rho \in S_\rho, \rho^* \in S_{\rho^*}, \sigma^2 = 1; k \in S_k \}, \end{aligned} \quad (2.11)$$

Each combination of parameters ω in the parameter space Ω leads to a model M_w , where $w = 1, \dots, W$. These set of values are chosen to cover the range of estimates obtained in most longitudinal studies (Verbeke and Molenberghs, 2000).

The ordering of the models for different combinations of parameters is illustrated in Table 2.1. In this table, the fastest varying index is the index for the autocorrelation parameter $\rho \in S_\rho$, whereas the slowest varying index is always the index for the random intercept variance $d_{11} \in S_{d_{11}}$. Notice that model complexity increases with the value of ω and that in the linear case ($p = 2$), the quadratic ($p = 3$) and cubic ($p = 4$) cases, ω belongs to the interval $[1, 216]$, $[217, 432]$ and $[433, 648]$, respectively. The ordering of models M_w starts with model class **AR1** and ends with **RI-RS- ρ^* -AR1**. The parameter combination of M_w can be illustrated by describing the easiest and most complex model classes as examples.

The first model class **AR1** ($M_1 - M_6$) is obtained when $k = 0$ and the ρ values are chosen one-by-one in ascending order from set S_ρ .

The last class of models **RI-RS- ρ^* -AR1** ($M_{97} - M_{216}$) is more complicated because the combination of parameters d_{11} , d_{22} and ρ^* is required while $k = 2$. So, let d_{11} , d_{22} and ρ^* be fixed at their lowest values, i.e., 1, 1 and 0.5, respectively. Next, choose one-by-one an autocorrelation parameter ρ from the set S_ρ , in ascending order. Afterwards, keep d_{11} and d_{22} unchanged and vary ρ^* to its higher level, i.e., 0.8. Then again, vary the autocorrelation parameter $\rho \in S_\rho$, in ascending order. Thereafter, let d_{11} remain at its lowest level, while d_{22} takes the value 3 and put the value of ρ^* back to its lowest level. Vary the autocorrelation parameter $\rho \in S_\rho$, in ascending order. Subsequently, let d_{11} and d_{22} be unchanged and increase ρ^* to 0.8. And, so again choose ρ from the set S_ρ . Repeat this procedure with a next d_{11} and so forth.

2.4.1. Algorithm

We numerically derive D_q -optimal designs ξ_q^* for all models M_w by optimizing the generalized variance of their estimated parameters. A program using function FMINSEARCH was implemented in MATLAB version 7.0.1(R14). The program uses a multidimensional unconstrained nonlinear minimization

Table 2.1: Ordering of linear mixed models M_w for the first-order polynomial model, $p = 2$, $q = 4$, $\sigma^2 = 1$, $w = 1, \dots, 216$.

M_w	Model Class	S_k	$S_{d_{11}}$	$S_{d_{22}}$	S_{ρ^*}	S_ρ
M_1-M_6	AR1	k=0: $D = \mathbf{0}$	—	—	—	$\left\{ \begin{array}{c} 0.0001 \\ 0.1 \\ 0.2 \\ 0.3 \\ 0.6 \\ 0.9 \end{array} \right\}$
M_7-M_{36}	RI-AR1	k=1: $D = (d_{11})$	$\{1, 2, 3, 4, 10\}$	—	—	$\left\{ \begin{array}{c} 0.0001 \\ 0.1 \\ 0.2 \\ 0.3 \\ 0.6 \\ 0.9 \end{array} \right\}$
$M_{37}-M_{96}$	RI-RS-AR1	k=2: $D = \text{Diag}(d_{11}, d_{22})$	$\{1, 2, 3, 4, 10\}$	$\left\{ \begin{array}{c} 1 \\ 3 \end{array} \right\}$	—	$\left\{ \begin{array}{c} 0.0001 \\ 0.1 \\ 0.2 \\ 0.3 \\ 0.6 \\ 0.9 \end{array} \right\}$
$M_{97}-M_{216}$	RI-RS- ρ^* -AR1	k=2: $D = \begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix}$ $d_{12} = \rho^* \sqrt{d_{11}d_{22}}$	$\{1, 2, 3, 4, 10\}$	$\left\{ \begin{array}{c} 1 \\ 3 \end{array} \right\}$	$\left\{ \begin{array}{c} 0.5 \\ 0.8 \end{array} \right\}$	$\left\{ \begin{array}{c} 0.0001 \\ 0.1 \\ 0.2 \\ 0.3 \\ 0.6 \\ 0.9 \end{array} \right\}$

algorithm developed by Nelder and Mead (Lagarias et.al., 1998; Matlab, 2004). The program requires an initial design value ξ_0 . We used equally-spaced designs as initial values.

Thereafter, we compute the RE in (2.9) for all D_q -optimal designs ξ_q^* 's. We evaluate $W \times W$ pairs of D_q -optimal true ξ_T^* and D_q -optimal selected ξ_S^* designs. The application of the criterion (2.10) leads to the maximin value MMV and the corresponding maximin design ξ_{MMV} . The total number of pairwise comparisons of designs and models is $648^2 - 648 = 419\,256$.

The MATLAB programs used to derive D_q -optimal designs and to find maximin designs in the parameter space Ω are available upon request to the authors as well as the list with the most highly efficient designs. The computational time required to obtain our results in a desktop computer with double INTEL processor *x86*, Family 6, Model 15 having a speed of 1595 MHz is approx. 35 minutes.

2.5. Results

A total of 648 minimum relative efficiencies are displayed in Figure 2.1, where the upper, middle and lower subplots show the minimum RE s for the locally optimal design per model, with a separate plot for (a) linear, (b) quadratic and (c) cubic models, respectively. On the x -axis of each subplot, we distinguished the four model classes: **AR1**, **RI-AR1**, **RI-RS-AR1**, and **RI-RS- ρ^* -AR1** represented by their optimal designs ξ_S^* where $S = 1, \dots, 648$.

In all subplots of Figure 2.1, a horizontal line is drawn at $\min RE = 0.85$ as reference line. This reference line leads to 97 optimal designs with $\min RE \geq 0.85$. So, at the design stage, assuming a cubic model ($q = 4$) is the best strategy against misspecification of the model and its variance covariance structure. Notice that all quadratic models and related optimal designs ξ_S^* turn out to have $\min RE$ values smaller than $RE = 0.85$ (Figure 2.1b).

The following results are based on Figure 2.1:

1. The efficiency of the encountered maximin design is very high

The maximin D_q -optimal design ξ_{MMV} for the parameter space Ω has a relative efficiency $RE = 0.8727$, see Figure 2.1a. The design

$\xi_{MMV} = \{-1 \quad -0.4568 \quad 0.4568 \quad 1\}$ is D_q -optimal for a first-degree polynomial with $p = 2$, random intercept $d_{11} = 1$ and with autocorrelation parameter $\rho = 0.2$.

Figure 2.1: Minima of relative efficiencies for D_q -optimal designs of families of linear mixed models

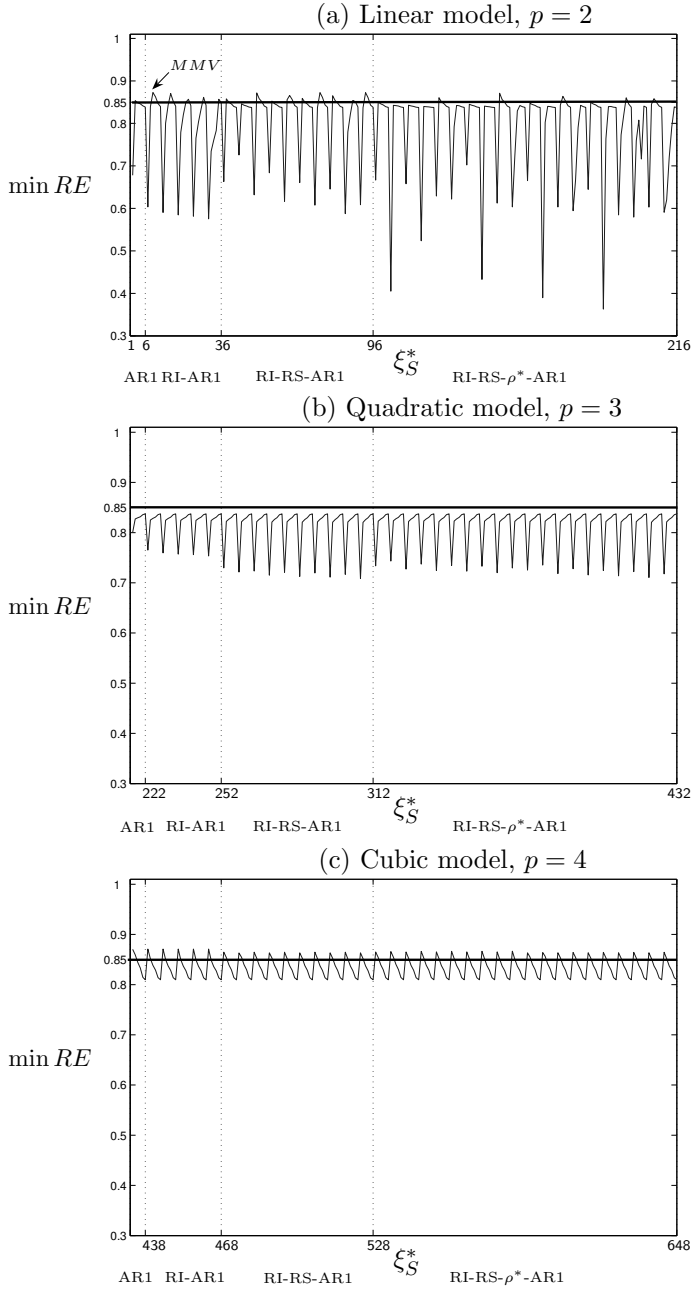


Table 2.2: Top 5: D_q -optimal designs with highest min RE 's

D_q -designs				min RE	p	ρ	D
{-1	-0.4568	0.4568	1}	0.8727	2	0.2	$\begin{bmatrix} 1 \end{bmatrix}$
{-1	-0.4567	0.4568	1}	0.8726	2	0.2	$\begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$
{-1	-0.4567	0.4568	1}	0.8726	2	0.2	$\begin{bmatrix} 10 & 0 \\ 0 & 3 \end{bmatrix}$
{-1	-0.4538	0.4538	1}	0.8718	2	0.1	$\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$
{-1	-0.4494	0.4494	1}	0.8710	4	0.0001	$\begin{bmatrix} 4 \end{bmatrix}$

2. The highly efficient designs are located around ξ^* for $q = 4$

In Table 2.2, a list of the most highly efficient D_q -optimal designs is presented. For practical purposes, these designs hardly differ from each other, i.e. the location of the time points is about the same and all D_q designs have an autocorrelation coefficient ρ smaller than 0.2. The models involved are mostly linear and pertain to either the class model **RI-AR1** or class model **RI-RS-AR1**.

3. Variation of minimum RE values decreases as the order of p increases

Figures 2.1a-2.1c show that the variation of the min RE values for the linear, quadratic and cubic models decreases as the degree of the polynomial increases.

The min RE s of locally optimal designs D_q shows steadily less variation on their min RE s values as the number of fixed parameters p increases from 2 to 4, i.e., as p approaches to number of time points q . We observe that under these conditions the locally optimal design points depend less on the covariance-variance structure and hence, are less affected by its incorrect specification.

4. Effect of autocorrelation parameter ρ on relative efficiencies

The effect of the autocorrelation parameter ρ in the relative efficiencies min RE of D_q -optimal designs corresponding to linear and quadratic models is large. For the linear model in Figure 2.1a, an increase of the value of ρ first leads to a rapid increase of the relative efficiency in the interval $[0.0001, 0.2]$. Once a maximum is reached in the interval $[0.1, 0.2]$, the relative efficiency of the D_q -optimal designs decreases slowly, but steadily, as correlation increases. All linear models studied

show this non-linear relation. For the quadratic model, the relation of $\min RE$ with the autocorrelation parameter ρ is positive. These $\min RE$'s reach their maximum values for $\rho = 0.9$, see Figure 2.1b.

In contrast to linear and quadratic models, the relation in cubic models between the autocorrelation parameter ρ and the relative efficiency $\min RE$ is relatively small and negative (see Figure 2.1c). High $\min RE$'s are usually encountered for the case where $\rho = 0.0001$ and low $\min RE$'s are found for high serial correlation values, i.e., $\rho = 0.9$.

2.6. Special cases

Effect of fixing design points to $q = 4$

The effect of fixing the number of design points to $q = 4$ in the maximin procedure is that the MMV of the maximin design will be very high. However, Wong (1994) and Moerbeek (2004) reported RE 's for uncorrelated fixed effects and fixed effects with AR1 structure, which are much lower. In their papers, the RE in equation (2.9) is evaluated for the $q = 2$ points D -optimal design $\xi^* = \{-1 \ 1\}$ and the $q = 3$ points D -optimal design $\xi^* = \{-1 \ 0 \ 1\}$ of a linear and quadratic model, respectively. Our results are based on D_q -optimal designs with fixed time points, $q = 4$. We will illustrate the consequences of such a choice.

Our maximin design is $\xi_{MMV} = \{-1 \ -0.4568 \ 0.4568 \ 1\}$ with $MMV = 0.8727$. If we compare the efficiency of ξ_{MMV} with that of the efficiency of that of $q = 2$ points ($\xi^* = \{-1 \ 1\}$) then the relative efficiency value is $RE = 0.4963$. This value is almost two times smaller than our current reported MMV . This difference is to be expected, because Wong (1994) and Moerbeek (2004) are comparing a D -optimal design with $q = 2$ points to a design with $q = 4$ design points, whereas we fix the design points to $q = 4$ points. We choose for the D_q -optimality criterion to account for the uncertainty about the order of the polynomial. Because, the effect of this additional uncertainty about the true order of the polynomial p_T adds extra loss of efficiency.

Effect of the presence of autocorrelation

In longitudinal and repeated measure designs, repeated measures could be correlated and in practice the uncorrelated condition in our numerical analysis might be unrealistic. So, let us introduce two sets $S_{\rho(1)}$, such that $S_{\rho(1)} \equiv S_{\rho}$ and $S_{\rho(2)} = \{0.1, 0.2, 0.3, 0.6, 0.9\}$.

In Table 2.3, the two cases are summarized. For $\rho \in S_{\rho(1)}$, we obtain a maximin design $\xi_{MMV} = \{-1 \ -0.4568 \ 0.4568 \ 1\}$ corresponding to a

Table 2.3: Maximin designs for uncorrelated and correlated cases

		ξ_{MMV}				MMV	p	ρ	D
$S_{\rho(1)}$	$\{-1$	-0.4568	0.4568	$1\}$		0.8727	2	0.2	$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$
$S_{\rho(2)}$	$\{-1$	-0.4247	0.4247	$1\}$		0.9734	2	0.3	$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

linear model with an AR1 structure having $\rho = 0.2$ and a random intercept variance $d_{11} = 1$. Its MMV is $RE = 0.8727$. For the $\rho \in S_{\rho(2)}$ condition, the maximin design becomes $\xi_{MMV} = \{-1 \ -0.4247 \ 0.4247 \ 1\}$, corresponding to a linear model with random intercept variance $d_{11} = 1$, $\rho = 0.3$ and its MMV has a $RE = 0.9734$. So, if we assume that the correlation $\rho \in S_{\rho(2)}$, our maximin procedure will lead to a maximin design with a much higher MMV .

The purpose of studying the effect of the autocorrelation on the chosen maximin design is to answer the question: what-if one has serial correlation. Do we need to know the value of the serial correlation parameter ρ to apply a maximin design? The answer to this question is: No. Once the practitioner knows if there is serial correlation, it does not seem to matter anymore what the value of ρ is; unless the follow-up time is very short.

Effect of using of equidistant time points

In practice, designs with equally-spaced time points are often used. In Table 2.4, the RE s of two equally-spaced designs are shown. These two locally optimal designs perform with min RE s of 0.8373 ($\rho \in S_{\rho(1)}$) and 0.9236 ($\rho \in S_{\rho(2)}$) when used in all possible model and variance covariance mis-specifications.

Table 2.4: Relative efficiencies RE for equidistant designs ξ_{equal} given uncorrelated and correlated cases

		ξ_{equal}				min RE	p	ρ	D
$S_{\rho(1)}$	$\{-1$	-0.3333	0.3333	$1\}$		0.8373	3	0.9	$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$
$S_{\rho(2)}$	$\{-1$	-0.3333	0.3333	$1\}$		0.9236	3	0.9	$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

Notice that the first min RE is slightly lower than the reference value 0.85. Moreover, it can be seen that the min RE increases when the autocorrelation parameter corresponding to the uncorrelated case is not considered in our parameter space.

2.6.1. Discussion on Hamilton Depression Scores' example

Hedeker and Gibbons reported the following estimates for a linear mixed model of first-degree ($p = 2$) with $q = 6$ design points: random intercept

variance $\hat{d}_{11} = 11.64$, random slope variance $\hat{d}_{22} = 2.08$, random covariance $\hat{d}_{12} = -1.40$ with $\hat{\sigma}^2 = 12.22$ and autocorrelation coefficient $\hat{\rho} = 0.37$ (Hedeker and Gibbons, 1996).

The generalized variance of the equally spaced designs $\xi_{\text{equal}} = \{-1 \ -0.6 \ -0.2 \ 0.2 \ 0.6 \ 1\}$ can be compared with that of the maximin design $\xi_{MMV} = \{-1 \ -0.4568 \ 0.4568 \ 1\}$ using the relative efficiency measure given in (2.9). Further, assuming that the model proposed in Hedeker and Gibbons (1996) is the true model, the ratio between these efficiencies is adjusted to account for the number of repeated measures of the designs ξ_{MMV} and ξ_{equal} , i.e. four and six, respectively. We find that the relative efficiency $RE(\xi_{MMV}|\xi_{\text{equal}})$ is 0.663, i.e., the equally spaced design should be replicated 1.5 times to have the same efficiency as our maximin design.

Thus, the maximin design proposed for the Hamilton's depression example is $\xi_{MMV} = \{0 \ 9 \ 26 \ 35\}$ days and the sample size connected to this design has 44 patients place at four time points instead of the 66 patients being measured at six time points.

2.7. Conclusions

In this paper, we have studied the optimality of designs with four types of incorrect specifications in the linear mixed models, namely the order of the polynomial $(p - 1)$, the size of the autocorrelation coefficient ρ , the number of random effects k , and the correlation ρ^* between the random intercept and random slope. Our results are limited to four-points designs for first-, second- and third-degree polynomials. However, the same methodology to find maximin designs can be applied to other misspecifications.

The application of the maximin criterion and the constraint $RE \geq 0.85$ has rendered 97 highly efficient designs. Our main conclusions are that, the maximin D_q -optimal design is highly efficient. The maximin D_q -optimal design $\xi_{MMV} = \{-1 \ -0.4568 \ 0.4568 \ 1\}$ has a relative efficiency $RE = 0.8727$ in the chosen parameter and model spaces. This maximin design is a D_q -optimal design for a first-degree polynomial having covariance structure given by a random intercept and an AR1 structure with autocorrelation $\rho = 0.2$. A list of top-five highest efficient designs that resembles the maximin design very much, is also provided. Moreover, it is shown that the variation of $\min RE$ values decreases as the order of $(p - 1)$ increases. Although equally-spaced designs are often used in practice, they are less efficient than the top-five highest efficient designs of our study. It means that when practitioners have budget restrictions or simply a small sample, the maximin designs proposed here could play a deciding role on choosing

a design over another and will perform better than equally-spaced designs. Equally-spaced designs will generally be the second best choice under model uncertainties, (cf. Tan and Berger, 1999).

Notice that due to computational constraints, we have chosen few parameter values (2 or 3) for the correlation parameter between the random intercept and random slope variances (ρ^*) and the random slope variance (d_{22}). From Figure 2.1(a) one can notice that the optimal designs for $p = 2$ depend strongly on the variance- covariance structure of their models. However, not all the parameters of the variance-covariance structure seem to affect the optimal designs and the relative efficiencies in the same manner. The largest changes on the optimal designs and *REs* are due to the value of the parameter ρ and the presence/absence of ρ^* and d_{22} rather than the values of the parameters themselves. So, we expect that the use of a finer mesh to discretise the variance-covariance parameters will not influence the results drastically.

Another topic of further research may be to construct highly efficient designs for polynomial models with covariates. Covariates may vary or not over time. It might also be matter of further research to study the effect of using other structures than the AR(1) on the obtained maximin design. Finally, other extensions may be to consider unbalanced data, missing values or drop-out processes.

CHAPTER 3

On the D –optimal designs of misspecified heterogeneous
linear mixed models ¹

¹*Submitted for publication*

Abstract

The responses of repeated measurements at different time points are usually correlated and have heterogeneous variances. Researchers often collect these responses using conventional equally-spaced designs or optimal designs for homogeneous data. The argument for this practice is that they do not know beforehand the data variance structure and that optimal designs for homogeneous responses are available in the literature. In this paper, the incorrect assumption of homogeneous error variances on the efficiency of parameter estimators in the linear mixed models is tackled. Optimal designs for such models with a heterogeneous error covariance structure are reported. These optimal designs show moderate to large shifts as compared with those for linear mixed models with a homogeneous error covariance structure. The relative efficiency is used to measure the effect of using an optimal design for a model with an assumed homogeneous error covariance structure instead of its heterogeneous counterpart. We show that optimal designs for homogeneous data remain highly efficient regardless the number of parameters, the degree of the polynomial, the serial correlation and the random variances and covariances, if the error variance structure follows a straight-line model. This is not always the case when the error variance follows a convex or a concave trajectory.

Keywords: Heterogeneous error variance; Linear mixed models; D -optimality; First-order autocorrelation; Relative efficiency

3.1. Introduction

Linear mixed models for repeated measure designs are used increasingly in medical, social and health science research. The corresponding data do not often hold the usual assumptions of independency and homogeneity, i.e. there is (auto)correlation within experimental units and the variances at different time points are often not constant. The incorrect specification of the variance and covariance structure is undesirable because it not only leads to biased and hence incorrect estimation of the variance of parameter estimates, but it also results in an incorrect allocation of the experimental units and thus in a suboptimal choice of the design.

Much of the literature on optimal designs for heteroscedastic models deals with independent errors. Researchers usually assume that the shape of the efficiency function of the error variance structure is (partially) known. See, for example Atkinson and Cook (1995) and Dette et.al. (2001). Another approach is to focus on cases when the heteroscedastic errors can be treated as if they were homoscedastic. See e.g. Wong and Cook (1993), Wong (1995) and Montepiedra and Wong (2001).

To our knowledge, the closest similar study to ours is a numerical study on D -optimal designs efficiency for polynomial regression of known or unknown degree and without serial correlation, which was reported by Antille and Weinberg (2000). They studied the efficiency loss of several functions of the error variance structure on the location of D -optimal points.

In this paper, we will study the effect of an incorrect specification of error variance structure on the efficiency of optimal designs for linear mixed models. The responses are modeled as polynomial functions of unknown regression parameters and the within subject covariances are described as functions of a set of known variance-covariance parameters. An autoregressive heterogeneous error covariance structure, ARH(1), is introduced to account for the heterogeneous error variability. D -optimal designs are found numerically. To measure efficiency loss, the relative efficiency (RE) of an optimal design for a model with an (incorrectly assumed) homogeneous autoregressive error covariance structure AR(1) is compared to the optimal design for a model with a (true) heterogeneous autoregressive error covariance structure ARH(1).

The outline of the paper is as follows. In section 3.3, we will introduce linear mixed models with a heterogeneous error variance structure, the corresponding D -optimal designs and relative efficiencies. In section 3.4, the classes of linear mixed models under study are explained. In section 3.5, the numerical results are presented. In section 3.6, we will discuss the results

and summarize our conclusions. However, first a motivating example from the field of health sciences is provided.

3.2. Longitudinal study of elderly chronic care populations

Modeling the decline of mental state of elderly chronic care populations is often crucial in large scale gerontological studies. Petkova and Teresi (2002) and Holmes et.al. (2000), for example, used the same data of a decade-long investigation of the impact of specialized care units for individuals with dementia. Petkova and Teresi (2002) compare the decline of quality of life and functional limitation over time of demented and non-demented elderly residents in long-term facilities. Functional limitations imply the capability to perform basic tasks e.g. eating, dressing and grooming. Several variables were measured at different occasions over an interval of several months and linear mixed models were used to describe the data. Such studies typically consider polynomial functions to describe trends over time. In connection to the application of linear mixed models several error covariance structures are possible, among which the autocorrelation structure is the most prominent for time-structured data. See, for example Holmes et.al. (2000).

Petkova and Teresi (2002) and Holmes et.al. (2000) also considered the case when variances of the outcome at different observation times were not the same. This case is referred to as the heterogeneous autoregressive error covariance structure. In order to estimate the polynomial parameters in a mixed effects model as efficiently as possible, the researcher should first need to optimally design the study. By doing this, it is usually assumed that the model has a homogeneous error covariance structure, with equal variances over time. An optimal design for such a homogeneous error covariance structure would then be the best choice. In many cases, however, the variances of the measurements often change over time and the data then show a heterogeneous error covariance structure (Petkova and Teresi, 2002). In this paper, the optimal designs for heterogeneous error covariance structure will be studied along with the relative efficiencies change caused by choosing incorrectly an optimal design for a homogeneous covariance structure.

3.3. Linear mixed model

Let \mathbf{y}_i denote a continuous observable response of subjects, $i = 1, \dots, n$. We assume that \mathbf{y}_i has three components: the fixed-effects $X_i\boldsymbol{\beta}$, the random-effects $Z_i\mathbf{b}_i$ and the error terms \mathbf{e}_i . The class of linear mixed-effects models

has the general form:

$$\mathbf{y}_i = X_i \boldsymbol{\beta} + Z_i \mathbf{b}_i + \mathbf{e}_i, \quad (3.1)$$

where $\boldsymbol{\beta}$ denotes a $p \times 1$ vector of unknown fixed parameters. X_i is a known $q \times p$ design matrix linking $\boldsymbol{\beta}$ to \mathbf{y}_i ; p and q represent the number of the fixed parameters of the polynomial regression and the number of time points, respectively. Let \mathbf{b}_i denote a $k \times 1$ vector of unknown random-effects parameters and Z_i be a known $q \times k$ design matrix linking \mathbf{b}_i to \mathbf{y}_i . \mathbf{b}_i represents how the i^{th} subject deviates from the average population and it is assumed to be normally distributed with mean $\mathbf{0}$ and $k \times k$ covariance matrix D . As illustration of model (3.1), the description of the decline of the functional limitations of the patients in the longitudinal study of Petkova and Teresi (2002) and Holmes et.al. (2000) is used, i.e. the fixed parameter vector $\boldsymbol{\beta}$ and the random parameter vector \mathbf{b}_i consist of polynomial parameters.

The term \mathbf{e}_i represents an unobservable error assumed to be normally distributed with mean $\mathbf{0}$ and covariance $\boldsymbol{\Sigma}_i^{1/2} \boldsymbol{\Psi}_i \boldsymbol{\Sigma}_i^{1/2}$. $\boldsymbol{\Sigma}_i$ is a matrix with variances $(\sigma_{1i}^2, \dots, \sigma_{qi}^2)$ on its main diagonal and $\boldsymbol{\Psi}_i$ is a positive-definite autocorrelation matrix depending on the parameter ρ . The parameter ρ denotes the autocorrelation between the j^{th} and j'^{th} time points x_j and $x_{j'}$, respectively. The elements of the autocorrelation matrix are given by $\psi_{jj'} = \rho^{\text{abs}(x_j - x_{j'})}$, where $j, j' = 1, 2, \dots, q$. These elements of the within-subject correlation matrix $\boldsymbol{\Psi}_i$ correspond to an exponentially decreasing function known as autoregressive correlation structure in which the time points x_j are not necessarily equally spaced. Note that \mathbf{b}_i and \mathbf{e}_i are independent from each other.

In health sciences and medicine, time points of repeated measurements studies are often planned to be the same for all individuals. See, as example, Petkova and Teresi (2002). In the following, we will assume that design matrices X_i and Z_i are the same for each subject and that the variances $\boldsymbol{\Sigma}_i^{1/2} \boldsymbol{\Psi}_i \boldsymbol{\Sigma}_i^{1/2}$ are the same for all subjects. Model (3.1) can now be written as $\mathbf{Y} = X\boldsymbol{\beta} + Z\mathbf{b} + \mathbf{e}$, where $\mathbf{Y}, X, \boldsymbol{\beta}, Z, \mathbf{b}$, and \mathbf{e} are all matrices concatenating n subject-specific regression models, i.e., $\mathbf{Y} = (\mathbf{y}_1^T, \dots, \mathbf{y}_n^T)^T$, $X = (X_1 | \dots | X_n)^T$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_p)^T$, $Z = \text{Diag}(Z_1, \dots, Z_n)$, $\mathbf{b} = (\mathbf{b}_1^T, \dots, \mathbf{b}_n^T)^T$ and $\mathbf{e} = (\mathbf{e}_1^T, \dots, \mathbf{e}_n^T)^T$. This linear mixed model implies a marginal model \mathbf{Y} that is normally distributed with mean $X\boldsymbol{\beta}$ and covariance matrix:

$$V = ZDZ^T + \boldsymbol{\Sigma}^{1/2} \boldsymbol{\Psi} \boldsymbol{\Sigma}^{1/2}, \quad (3.2)$$

where $D = \text{Diag}(D, \dots, D)$, $\boldsymbol{\Psi} = \text{Diag}(\boldsymbol{\Psi}_1, \dots, \boldsymbol{\Psi}_n)$ and $\boldsymbol{\Sigma}^{1/2} = \text{Diag}(\boldsymbol{\Sigma}_1^{1/2}, \dots, \boldsymbol{\Sigma}_n^{1/2})$.

$\dots, \Sigma_n^{1/2}$). The matrix V has size $(nq \times nq)$. Notice that the diagonal matrix Σ for the linear mixed models is similar to the efficiency function given by Fang et.al. (2006). In this paper, two cases for the diagonal matrix Σ will be considered. A constant error variance function (i.e. $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_q^2 = \sigma^2$) so that the homogeneous error covariance structure reduces from $\Sigma^{1/2}\Psi\Sigma^{1/2}$ to $\sigma^2\Psi$. This case will be referred to as AR1. The second case is a non-constant error variance function leading to a heterogeneous error covariance structure $\Sigma^{1/2}\Psi\Sigma^{1/2}$, where the diagonal elements of Σ are not equal in value. This case is referred to as ARH1.

The asymptotic covariance matrix of the best linear unbiased estimator $\hat{\beta}$ is proportional to:

$$\text{Var}(\hat{\beta}) = (X^T V^{-1} X)^{-1}, \quad (3.3)$$

which is equal to the inverse Fisher information matrix, i.e. small valued $\text{Var}(\hat{\beta})$ correspond to a large valued information matrix.

3.3.1. Optimal designs

A continuous design ξ for a linear mixed models can be defined as:

$$\xi = \left\{ \begin{array}{cccccc} x_1 & x_2 & \dots & x_j & \dots & x_q \\ w_1 & w_2 & \dots & w_j & \dots & w_q \end{array} \right\}, \quad \text{where } j = 1, 2, \dots, q. \quad (3.4)$$

The values of x_j are confined to a given design space $\mathcal{X} = [-1, 1]$ and $\int_{\mathcal{X}} \xi(dx) = 1$. The weights w_j are in the interval $0 \leq w_j \leq 1$. The implemented design has Nw_j observations at x_j , $j = 1, 2, \dots, q$, where N is the total number of observations. See Atkinson and Donev (1996) among others. In longitudinal studies, balanced designs are usually used for modeling purposes, i.e. designs with $w_j = 1/q$, for all j .

The problem is to search for the D -optimal design ξ^* that minimizes the function $\text{Det}[\text{Var}(\hat{\beta})]$ over the design space \mathcal{X} and which satisfies:

$$\text{Det}[\text{Var}_M(\hat{\beta}|\xi^*)] \leq \text{Det}[\text{Var}_M(\hat{\beta}|\xi)] \quad \forall \xi \in \mathcal{X}, \quad (3.5)$$

where $\text{Var}_M(\hat{\beta}|\xi)$ is the variance-covariance matrix of $\hat{\beta}$ for a design ξ under model M in equation (3.1). The $\text{Det}[\text{Var}_M(\hat{\beta}|\xi)]$ is often referred to as the generalized variance of $\hat{\beta}$.

3.3.2. Measure of efficiency change

Two types of error covariance structures $\Sigma^{1/2}\Psi\Sigma^{1/2}$ that influence the D -optimal design ξ^* can be distinguished:

1. a homogeneous structure Σ_{hom} for an assumed model M_{hom} ; and
2. a heterogeneous structure Σ_{het} for the encountered model M_{het} .

M_{hom} is a model with homogeneous error variance-covariance structure $\Sigma_{\text{hom}}^{1/2}\Psi\Sigma_{\text{hom}}^{1/2} = \sigma^2\Psi$ and variance-covariance $V_{\text{hom}} = ZDZ^T + \sigma^2\Psi$ for the responses \mathbf{Y} ; while M_{het} is the model with heterogenous error variance-covariance structure $\Sigma_{\text{het}}^{1/2}\Psi\Sigma_{\text{het}}^{1/2}$ and variance-covariance $V_{\text{het}} = ZDZ^T + \Sigma_{\text{het}}^{1/2}\Psi\Sigma_{\text{het}}^{1/2}$ for the responses \mathbf{Y} , that fits the data best.

For the D -optimality criterion the relative efficiency is defined as:

$$RE = \left[\frac{\text{Det}\{\text{Var}(\hat{\beta}^*|\xi_{\text{het}})\}}{\text{Det}\{\text{Var}(\hat{\beta}^*|\xi_{\text{hom}})\}} \right]^{\frac{1}{p}}. \quad (3.6)$$

This ratio measures the efficiency change when a study is planned with an optimal design assuming an homogeneous error covariance structure for the data, while instead a heterogeneous error covariance structure should have been used. Notice that the number of the fixed parameters of the models is p and that the power $1/p$ results in an efficiency measure which is proportional to design size, irrespective of the dimension of the model. See Atkinson and Donev (1996) and Pukelsheim (1993) among others. The relative efficiency of, say, $RE = 0.6$ means that $(1/RE - 1) \times 100\%$ more observations are needed for the homogeneous optimal design to be as efficient as the true heterogenous optimal design.

3.4. Classes of linear mixed models

For our numerical analysis, we have distinguished four classes of linear mixed models based on different combinations of the parameter values. The values of the parameters were selected in such a way that they cover many cases found in practice. The responses are modeled as polynomial functions of unknown regression parameters given by (3.1). D -optimal designs for polynomials having as many design points as fixed model parameters, i.e. $q = p$ have been reported by Bunke and Bunke (1986, pp. 546), among others. In this paper, we also deal with D_q optimal designs, i.e. D -optimal designs where the number of time points q is fixed and the number of re-

gression parameters p is smaller than q . For further details on these optimal designs, the reader is referred to Tan and Berger (1999).

We will restrict ourselves to three polynomial models, namely the first-degree ($p = 2$, linear), second-degree ($p = 3$, quadratic) and the third-degree ($p = 4$, cubic). The reason for this restriction is that, in practice higher (fourth- or fifth-) degree polynomials are seldomly encountered in longitudinal studies in medical, social and health science research.

Ten different values for the autocorrelation coefficient ρ are considered in the interval $0 < \rho < 1$. With these values different autocorrelation matrices Ψ can be specified.

Class 1: Fixed-effect models

The variance-covariance matrix V for fixed-effect models in equation (3.2) reduces to $V = \Sigma^{1/2}\Psi\Sigma^{1/2}$, since $D = \mathbf{0}$, implying that the block-matrix D is zero as well. In total, 10 fixed-effect models, each with a different value of ρ were considered, assuming either a homogeneous (AR1) error structure or a heterogeneous error structure (ARH1).

Class 2: Random-intercept models

For random intercept models the variance-covariance matrix V in equation (3.2) reduces to $V = d_{11} ZZ^T + \Sigma^{1/2}\Psi\Sigma^{1/2}$. D is a scalar d_{11} representing the variance of the random intercept. The product ZZ^T is a blocked diagonal matrix with main diagonal $\mathbf{1}\mathbf{1}^T$, where $\mathbf{1}$ is a column vector of ones. The following values of d_{11} are distinguished: $\{1, 2, 3, 4, 10\}$. In total, 50 different random-intercept models, obtained for each combination of values of d_{11} and ρ , have been considered.

Class 3: Independent random-intercept and random-slope models

For random-intercept and random-slope models, the variance-covariance matrix D has the form $D = \text{Diag}(d_{11}, d_{22})$, where the covariance between the random-intercept and the random-slope is assumed to be zero. The AR1 or ARH1 correlation structures of the errors are represented by $\Sigma^{1/2}\Psi\Sigma^{1/2}$. Apart of the ten different values for ρ and five different values for d_{11} , two values for d_{22} are distinguished, namely $\{1, 3\}$. In total 100 different combinations of ρ , d_{11} and d_{22} distinguish 100 different random-intercept and random-slope models.

Class 4: Correlated random-intercept and random-slope models

The random-intercept and random-slope models in this class have two different correlations represented by ρ^* and ρ . The covariance matrix D of the random-effects with diagonal elements d_{11} and d_{22} and off-diagonal $d_{12} = d_{21}$. These off-diagonal elements are computed by $d_{12} = \rho^*\sqrt{d_{11}d_{22}}$, where ρ^* has two values $\{0.5, 0.8\}$. The second correlation is the autocorrelation parameter ρ in Ψ , with the already mentioned ten different values. The

combination of values of ρ, d_{11}, d_{22} and ρ^* leads to a total of 200 different models in this class.

In summary, in model classes 1, 2, 3, and 4, there are 10, 50, 100 and 200 different models, respectively. In total, there are $4 \times 360 = 1440$ different models (with AR1 or ARH1 structures) for each feasible combination of $p = 2, 3, 4$ and $q = 3, 4$. Notice that the total number of observations is kept constant (N) and that a description of the algorithm used in the numerical analysis is given in the appendix.

3.4.1. Homogeneous and heterogeneous error covariance structure

Within subject variation may be different at the design (time) points (x_j). That is data can have either a homogeneous (AR1) or a heterogeneous (ARH1) error variance structure Σ . The diagonal elements $\Sigma_{jj}(j = 1, \dots, q)$ of this matrix can be modeled by distinct functions. Here, four functions are used, i.e. a constant, a linear, a convex and a concave function. The last three functions correspond to a heterogeneous error variance, ARH1. These three heterogeneous structures cover a wide range of error variance functional matrices that are found in practice. These functions are displayed in Figure 3.1, and describe the increase of the error variances over the design interval $[-1, 1]$.

3.5. Results

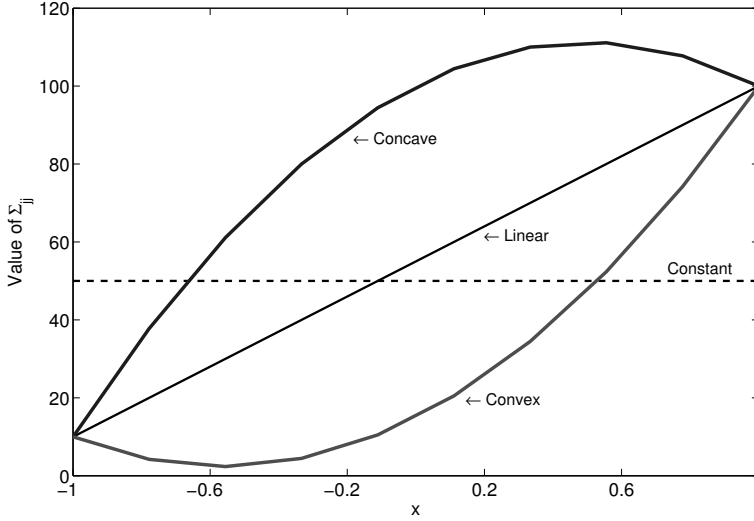
In this section, the D -optimal design points for linear mixed models with heterogeneous error variances are discussed along with their changes of efficiency when compared with their homogeneous counterparts.

3.5.1. D -optimal designs of random intercept models

The most striking effect of having heterogeneous errors in the variance-covariance structure of linear mixed models is that the corresponding optimal design points tends to shift in the direction where the error variances are smaller. As an illustration of this effect, the D -optimal design points for a fixed-effects model (class 1) and a random-intercept model(class 2) of first-, second- and third-degree polynomial models for $q = p = 3$ and $q = p = 4$ are displayed in Figures 3.2 and 3.3, respectively.

In these two figures, the D - optimal design points for the constant, convex, linear, and concave error variance structures are displayed in the first-, second-, third- and fourth-column plots, respectively. In all plots the x -axis represents the design space $[-1, 1]$ and the y -axis corresponds to the

Figure 3.1: Heterogeneous variance functions



autocorrelation parameter ρ . In the plots of the upper-row of Figure 3.2, it can be seen that, the shift of the middle design point of a fixed-effects models is related to the form of the error variance structure. This shift is towards the left where the variances are smaller. The plots of the middle- and lower-row of the Figure 3.2 illustrate that when $q = p$ the middle point of the optimal design for the random-intercept model with linear error variance structure seems to be robust with respect to the value of the random intercept variance d_{11} . It should be noted that D -optimal designs for model classes 3 and 4 are omitted since these plots are very similar to those for model class 2. Moreover, it is worth mentioning that when $p < q$ the optimal design points (not shown here) can shift dramatically. We come back to this point in section 3.5.2. Notice also that the results of the first-degree polynomial are omitted. The reason is that D -optimal designs of linear mixed models for a first-degree polynomial remain unchanged, since their two design points are fixed at the endpoints of the design region $[-1, 1]$.

3.5.2. Efficiency loss

To measure efficiency loss, the relative efficiency (RE) of an optimal design for a model with an (incorrectly assumed) homogeneous autoregressive

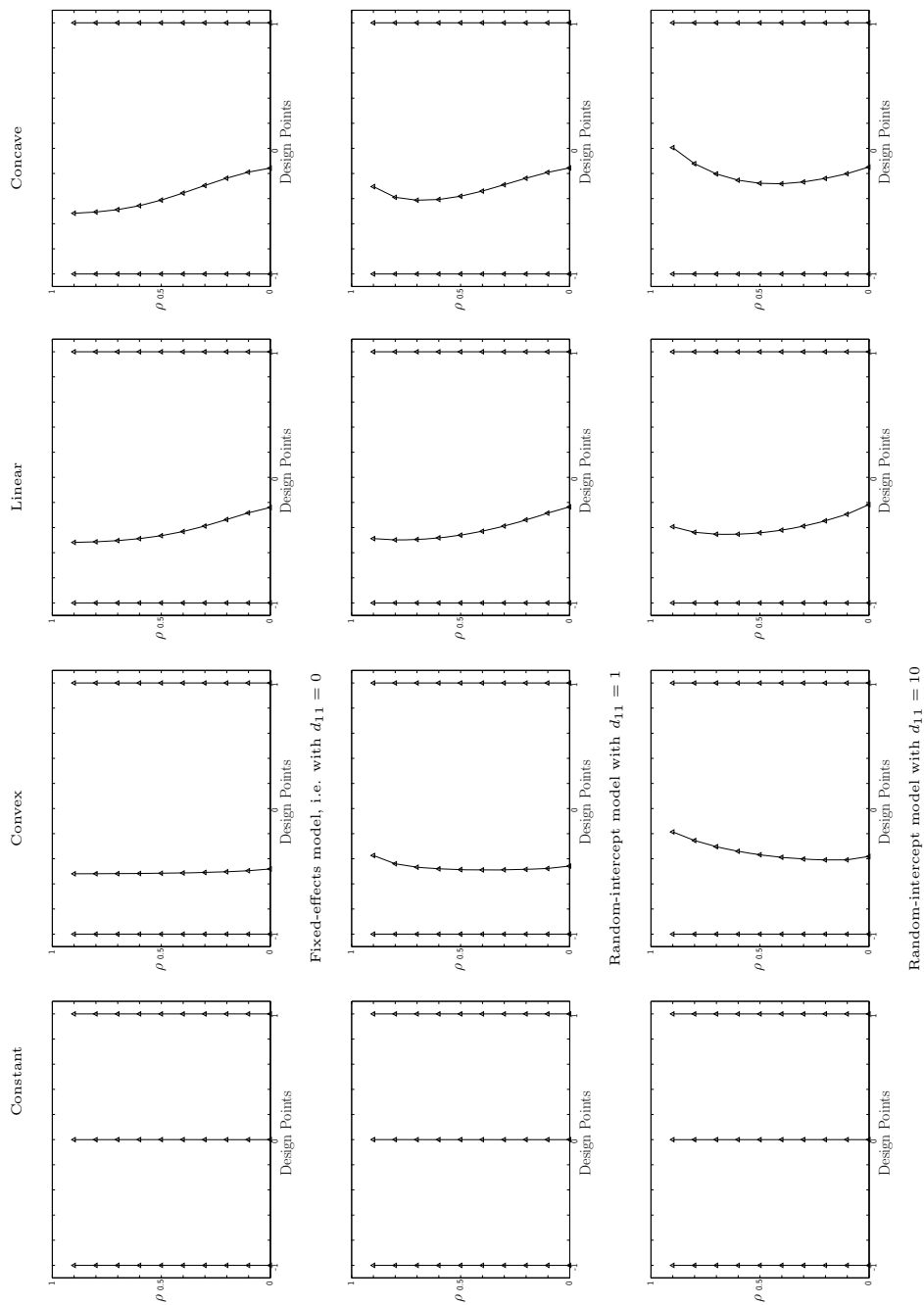
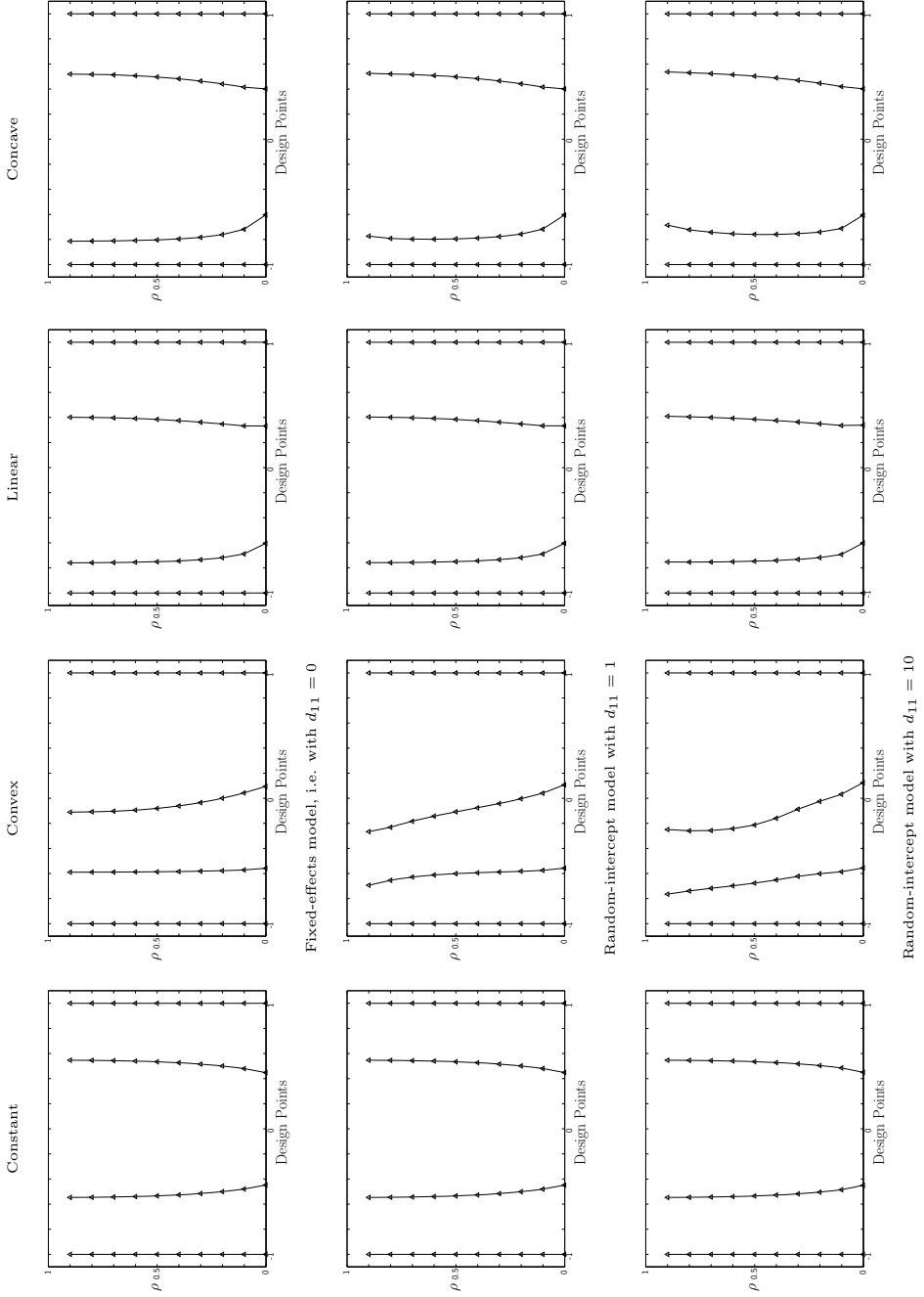
Figure 3.2: D -optimal design points for heterogeneous models of second-degree $p = 3$ 

Figure 3.3: D -optimal design points for heterogeneous models of third-degree $p = 4$ 

error covariance structure $AR(1)$ is compared to the optimal design for a model with a (true) heterogeneous autoregressive error covariance structure $ARH(1)$.

Without loss of generality, we will mainly describe the results of model classes 1 and 2, i.e., when $d_{11} = 0, 1$ (and $d_{22} = \rho^* = 0$) for the first-, second- and third-degree polynomial models connected to the design (time) points $q = 3, 4$ and $\rho = 0, 0.2, 0.4, 0.6, 0.8$. Table 3.1 displays the RE 's of 50 (out of the 360) models introduced in section 3.4.

The values under the columns (labeled with) Convex, Linear and Concave correspond to the relative efficiencies when D -optimal designs for fixed-effects models (class 1) and random-intercept models (class 2) with assumed homogeneous error variance are used instead of those optimal designs for similar models but with a heterogeneous error variance structure of convex, linear and concave forms, respectively. The RE values of, for example the first five rows, show the relative efficiencies for the first-order polynomial model ($p = 2$) with fixed-effects ($d_{11} = 0$) in the interval $0 < \rho \leq 0.8$ for the case of $q = 3$ and 4, respectively.

By looking at Table 3.1, the following conclusions can be drawn:

1. The linear error variance structure leads to high relative efficiency values, regardless whether the number of designs points is equal or larger than the number of parameters, i.e. $RE > 0.91$ when $q \geq p$.
2. The concave error variance structure does not lead to large losses of efficiency when the number of designs points is equal to the number of parameters, i.e. $RE > 0.93$ when $q = p$.
3. However, in other cases, say, when $q > p$ and the error variance structure is convex or concave, large to moderate losses of efficiency can occur. Moreover, these losses do change, rather strongly, based on the value of the serial correlation. See for example the designs with three time points ($q = 3$) for highly correlated ($\rho = 0.8$) linear ($p = 2$) fixed-effects models. The convex- and concave-related efficiency losses are up to 62% and up to 30%, respectively.

We have no other intuitive explanation to the question why the relative efficiencies of the convex error variance structure are lower than those of the concave but that it has to do with the location of their corresponding optimal design points. Notice that the optimal design points with a convex error variance structure show the largest shift. Compare for example, the location of the design time points provided in the plots of the first, second

Table 3.1: Relative efficiencies for the first-, second- and third-degree polynomial models.

		Convex	Linear $q = 3$	Concave	Convex	Linear $q = 4$	Concave
$p = 2$	$d_{11}=0$	0	0.6003	0.9996	0.9984	0.9292	0.9748
		0.2	0.6598	0.9983	0.9989	0.8172	0.9997
		0.4	0.5804	0.9921	0.8517	0.7371	0.9817
		0.6	0.4872	0.9514	0.7791	0.6693	0.9587
		0.8	0.3843	0.9133	0.7037	0.6092	0.9288
	$d_{11} = 1$	0	0.6911	0.9996	0.9986	0.9479	0.9679
		0.2	0.7335	0.9984	0.9991	0.8616	0.9997
		0.4	0.6574	0.9931	0.8663	0.7951	0.9832
		0.6	0.5784	0.9563	0.8043	0.7522	0.9628
		0.8	0.5184	0.9249	0.7532	0.7463	0.9390
	$d_{11} = 0$	0	0.6651	0.9672	0.9883	0.8050	0.9811
		0.2	0.6341	0.9545	0.9825	0.8072	0.9680
		0.4	0.6172	0.9401	0.9739	0.7758	0.9540
		0.6	0.6088	0.9299	0.9657	0.7438	0.9467
		0.8	0.6053	0.9248	0.9608	0.6936	0.9420
$p = 3$	$d_{11} = 1$	0	0.7169	0.9684	0.9885	0.8460	0.9826
		0.2	0.6785	0.9541	0.9824	0.8479	0.9678
		0.4	0.6675	0.9405	0.9754	0.8234	0.9548
		0.6	0.6763	0.9319	0.9713	0.8096	0.9491
		0.8	0.7210	0.9303	0.9750	0.8017	0.9479
	$d_{11} = 0$	0				0.8846	0.9650
		0.2				0.8882	0.9475
		0.4				0.8866	0.9450
		0.6				0.8841	0.9445
		0.8				0.8826	0.9445
$p = 4$	$d_{11} = 1$	0				0.8982	0.9658
		0.2				0.8919	0.9476
		0.4				0.8906	0.9454
		0.6				0.8901	0.9454
		0.8				0.8876	0.9462

and third columns of Figure 3.2 and 3.3. These shifts are the largest for the convex error variance structure and the least for the linear error variance.

3.5.3. Consequences for the longitudinal study of elderly chronic care populations

In the longitudinal study of Holmes et.al. (2000) and Petkova and Teresi (2002), three equally-spaced time points were considered. The planned interval among baseline, first follow-up and second follow-up was 6 months. The most suitable model was a linear model with serially correlated heterogeneous error variance. The design parameters are summarized as follow $\xi = \{0 \ 6 \ 12\}$, $q = 3$, $p = 2$, $\rho = 0.67$ and sample size $N = 186$.

Looking back to this example, three practical questions are put forward:

1. *Was the number and location of the design points optimal?*

Since the data analyzed by Petkova and Teresi (2002) is best fitted by a linear ($p = 2$) function, the optimal number of design points is two points, i.e. at the beginning and end of the design space. It has been shown (Tan and Berger, 1999) that for longitudinal studies with unknown random variance structure, the best strategy is to use the same number of time points as there are fixed parameters in the linear mixed model. The RE s for $p = q$ of Table 3.1 show that previous result also holds for optimal designs for linear mixed models with linear and concave heterogeneous error variance structure.

2. *How efficient is the equally-spaced design with three design points when compared with its heterogeneous counterpart?*

Based on the values reported in Table 3.1, it can be seen that the relative efficiency of using an equally-spaced (homogeneous) design (instead of that for a model with heterogeneous error variance structure) will change strongly based on the form of the heterogeneity, i.e. the value of the relative efficiency will be high ($RE = 0.95$) to moderate ($RE = 0.78$) when the error function is linear or concave, respectively. A low efficiency ($RE = 0.49$) is observed when the error function is convex. Consequently, the efficiency of a chosen design can be significantly improved if background information with regard to the direction and trajectory of the error variance structure is (partially) known.

3. *How does the most suitable optimal design look like and how does it affect the sample size?*

When the error variance structure is a concave, a linear or a convex function, the optimal designs are $\xi_{\text{het,Convex}}^* = \{0 \ 3 \ 12\}$, $\xi_{\text{het,Linear}}^* = \{0 \ 2 \ 12\}$ and $\xi_{\text{het,Concave}}^* = \{0 \ 1 \ 12\}$ months, respectively. Hence, if the equally spaced design $\xi = \{0 \ 6 \ 12\}$, is applied, as many as, 194, 10 and 52 extra patients are needed to have this homogeneous design as efficient as its heterogeneous convex, linear or concave optimal designs, respectively. This also means that if the longitudinal data follows a convex error structure, the optimal design $\xi_{\text{het,Convex}}^*$ will need about the half of the original 186 patients.

3.6. Conclusions

The purpose of this paper is to provide insight on the robustness of D -optimal designs with respect to the error covariance structure of the model.

Our main conclusions can be summarized as follows:

D -optimal designs for linear mixed models with an ARH1 error covariance structure show moderate to large shifts of their middle design points when compared with the optimal designs for linear mixed models with an AR1 error covariance structure.

The D -optimal designs for heterogeneous linear mixed models with a linear error variance function seem (relatively) unchangeable with respect to the number of fixed parameters, number of design points and the value of the serial correlation parameter.

Using a D -optimal design for a homogeneous linear mixed model instead of that for a heterogeneous model does not lead to major losses of efficiency when the error variance covariance is linear with $q \geq p$ or concave with $q = p$. High efficiency losses can be expected when the error variance is convex. In other words, the use of a suboptimal design, when the error variance is convex, can lead to considerable overestimation of the optimal sample size. So, when the researcher knows beforehand that the error variance structure is not convex, the equally-spaced designs are (still) highly efficient options.

Regular patterns were found over the different classes of models and RE values do not show additional effects with respect to the value of the random slope variance and correlation between random parameters. Furthermore, several convex and concave functions have been studied and the results are similar to those reported here. This may give some confidence to claim that our results also apply to other classes of models and error function structures.

Our main recommendation to practitioners is to try to acquire prior information on the underlying polynomial process and the error variance

profile before they choose an optimal design to collect longitudinal data.

Future research is encouraged to study the robustness of heterogeneous optimal designs under other optimality criteria.

Appendix: Algorithm

D -optimal designs in the compact interval $[-1, 1]$ are found numerically by optimizing the generalized variance of the parameter estimators. The program uses function `FMINCON` of MATLAB version 7.0.1(R14). This function is multidimensional constrained nonlinear minimization algorithms developed by Han & Powell. See Matlab (2004). Figures 3.2-3.3 were obtained using these algorithms.

We compute the relative efficiency RE given by (3.6) for each AR1 with respect to one of the three ARH1 structures. In this step, we compare one-to-one an optimal design for a homogeneous variance-covariance structure V_{hom} when instead a design for a heterogeneous variance-covariance structure, V_{het} for the responses in \mathbf{Y}_{het} should be used. A total of 5400 paired comparisons are executed, i.e., 2160 comparisons for the first-, same 2160 for the second-, and 1080 for the third-degree polynomial. An abstract of these results is provided in Table 3.1.

Our numerical results showed that some error variance functions may lead to singular values, i.e. causing unrestricted maximum likelihood estimation to break down. In the literature this phenomenon was mentioned by Crisp and Burridge (1994). They claimed that the exponential function does behave better and avoids this difficulty. We have studied the cases when this singularity may arise, i.e. $q = 3$ and $p = 2$. We observed that singularities arise when the intercept of the functions is zero. This might (implicitly) explain why by using exponential functions Crisp and Burridge (1994) avoided singularity.

CHAPTER 4

The effect of dropout on the efficiency of D –optimal designs
of linear mixed models ¹

¹*Statistics in Medicine*(2008) **27**:14, 2601–2617.

Abstract

Dropout is often encountered in longitudinal data. Optimal designs will usually not remain optimal in the presence of dropout. In this paper, we study D -optimal designs for linear mixed models where dropout is encountered. Moreover, we estimate the efficiency loss in cases where a D -optimal design for complete data is chosen instead of that for data with dropout. Two types of monotonically decreasing response probability functions are investigated to describe dropout. Our results show that the location of D -optimal design points for the dropout case will shift with respect to that for the complete and uncorrelated data case. Due to this shift, the information collected at the D -optimal design points for the complete data case does not correspond to the smallest variance. We show that the size of the displacement of the time points depends on the linear mixed model and that the efficiency loss is moderate.

Keywords: D -optimality, Dropout, Linear mixed models, Relative efficiency

4.1. Introduction

Dropout is a prominent problem in most longitudinal studies. Especially in longitudinal studies dealing with severe and invasive therapies, dropout is a frequently encountered pattern of missing data. Another field where dropout is unavoidable concerns the follow-up of clinical trials of disorders like depression and substance abuse.

In designing a longitudinal study, a researcher would want to choose an optimal design for the corresponding model. If, however, dropout is anticipated, the optimal design will be sub-optimal. Since the actual found dropout pattern will not be known in the design stage, it is important to investigate how much efficiency is lost if the optimal design for complete data is used when the data contains dropout. This loss of efficiency will be studied in this paper.

Optimal designs for the linear models have been previously published (Rao, 1973; Pukelsheim, 1993). Research on comparing the performance of fixed designs and optimal designs is been published by Tan and Berger (1999). They showed that fixed designs are in general less efficient than optimal designs. Especially when the number of time points is much larger than the number of parameters of the underlying linear model, the efficiency of optimal spaced designs will surpass that of fixed designs by forty to sixty percent. Recently Tekle, Tan and Berger (2008a) have also shown that optimal designs are more cost-effective than fixed designs.

Optimal designs, however, assume that the data over the repeated measures are complete, i.e., that there is no dropout. Optimal designs that do take dropout into account have been studied in Herzberg and Andrews (1976a,b); Hackl (1995); Imhof et.al. (2002) and Imhof et.al. (2004). Imhof et.al. (2002) provided closed-form maximin D -optimal designs for partially specified response probability functions and Imhof et.al. (2004) showed for a straight-line regression that the D -optimal designs for the complete data case remain efficient whenever the probabilities of missingness across the design space are not too different. They also found that the D -optimal design for the incomplete data case may lead to a substantial loss of efficiency when the dropout is extreme, i.e., 80%. These studies, however, are restricted to linear fixed-effects models with uncorrelated data.

A particularly popular model in longitudinal studies with continuous data is the linear mixed effects model. This model assumes that changes in time for each subject can be described by a linear regression model with fixed and random parameters along with possible serially correlated errors. The parameters in this model can be estimated by using maximum likelihood

estimation methods. This model and estimation methods can also facilitate studies with dropout in the data, see for example Verbeke and Lesaffre (1999).

Optimal designs for the linear mixed effect models have been previously published (Tan and Berger, 1999; Ouwens et.al., 2002) and the impact of unanticipated dropout as well as ways to allow dropout in linear mixed models at a design stage were reported by Galbraith and Marschner (2002). They found for a first-degree polynomial mixed effects model with a dropout percentage of no more than 30% that, the loss of power is not affected much by the choice of the number of design points.

At the design stage of longitudinal studies, researchers cannot easily determine whether potential dropout processes arise in their studies. The key questions to be solved are therefore (1) Does this expected dropout influence the relative efficiency of their optimal designs? (2) Should they account for this potential dropout? In this paper, we will investigate the effect of dropout on the efficiency of D -optimal designs of linear mixed models for longitudinal data by extending previous results. In section 4.3, we will describe the linear mixed model with a response probability function that incorporates dropout. In section 4.4, D -optimal designs for longitudinal data with dropout are presented. Optimal designs will be compared by means of their relative efficiencies. In other words, we compare the generalized variance of the estimated regression coefficients for a D -optimal design based on longitudinal data with dropout against that for the D -optimal design under the incorrect assumption of complete data; the latter generalized variance being adjusted by the underlying response probability function. Finally, section 4.5 contains a summary of the main conclusions. First, however, a motivational example of longitudinal data with loss to follow-up will be described in the next section.

4.2. Illustration: Randomized clinical trial

Complete data sets are rare in follow-up studies of clinical trials. Intensive efforts of obtaining high-rates of follow-up and efforts on tracking patients might sometimes yield meager results. The data from a 1-year follow-up of a randomized clinical trial of psychotherapy (cognitive-behavioral relapse prevention or clinical management) and pharmacotherapy (desipramine or placebo) for ambulatory cocaine addicts is taken as example (Carroll et.al., 1994; Nich et.al., 1997).

Patients were randomly assigned to one of four treatments. Outcomes were measured using urine toxicology screens and the Addiction Severity

Table 4.1: Dropout pattern of the psycho-and pharmacotherapy follow-up study for cocaine addicts (Carroll et.al., 1994).

	months				
	0	1	3	6	12
Four Interviews		25	25	25	25
Three Interviews		25	25	25	
Two Interviews		16	16		
One Interview		31			
Subjects (N)	121	97	66	50	25
index(j)	1	2	3	4	5
Response probability $p(x_j)$	1	0.80	0.54	0.41	0.21

Index. Patients were scheduled to be followed 1, 3, 6 and 12 months after termination of the study treatment. A total of 121 patients were included. The first month after treatment, the response percentage was 80%. Only 54% of the patients were measured at 1 and 3 months; 41% were measured at 1, 3 and 6 months; and as much as 21% of the subjects finished the follow-up. This dropout process is described in Table 4.1 in terms of a response probability function and the number of patients that are available at each design point.

The problem of finding an optimal design for linear mixed models with dropout is hampered by the fact that the response probability function is unknown at the design stage of a study. Although researchers may have reasonable guesses about this function, usually they do not know its exact form. The question arises, how much efficiency is lost if there is dropout and an optimal design that incorrectly assumed complete data is used to design the study, instead of the optimal design that is based in the unknown dropout process. In this paper, this problem will be studied.

In the next section, we will put forward a linear mixed model that incorporates a dropout process via a response probability function.

4.3. Linear mixed models

Linear mixed models have been discussed extensively, see e.g., Diggle et al. (1994); Verbeke and Molenberghs (2000). In these models, the response \mathbf{y}_i of an i^{th} subject is given by:

$$\mathbf{y}_i = X_i\boldsymbol{\beta} + Z_i\mathbf{b}_i + \mathbf{e}_i, \quad (4.1)$$

where β denotes a $p \times 1$ vector of parameters, X_i is a known $q \times p$ design matrix linking β to \mathbf{y}_i ; p and q represent the number of the fixed parameters of a polynomial regression and the number of design points, respectively. Let \mathbf{b}_i denote a $k \times 1$ vector of unknown subject-specific effects and Z_i be a known $q \times k$ design matrix linking \mathbf{b}_i to \mathbf{y}_i . The random-effects parameter vector \mathbf{b}_i represents how the i^{th} subject deviates from the average population, $\mathbf{b}_i \sim N(\mathbf{0}, D)$ and D is the variance-covariance matrix of the random-effects.

The errors \mathbf{e}_i are distributed as $N(\mathbf{0}, \sigma^2 \Psi_i)$. Here $\sigma^2 \Psi_i$ is a positive-definite covariance matrix and it depends on the correlation parameter ρ of the time lag between the j^{th} and j'^{th} time points x_j and $x_{j'}$. The elements of the within-subject correlation matrix Ψ_i follow a first-order autoregressive correlation, AR(1) with $\Psi_i(x_j, x_{j'}) = \rho^{\text{abs}(x_j - x_{j'})}$. We will restrict ourselves to this correlation structure since it is a very often encountered structure in longitudinal designs. For an overview of covariance structures and serial correlation structures, see Verbeke and Molenberghs (2000, p.99). We also assume that the random variables \mathbf{b}_i and \mathbf{e}_i are independent, identically and normally distributed.

In this article, we obtain the optimal allocation of design points x_j of longitudinal observations \mathbf{Y} for a sample of subjects, in cases where some of them might dropout at specific time points. In the next sections, a design ξ is introduced along with a response probability function and a reformulation of the linear mixed model to handle longitudinal data with dropouts.

4.3.1. Design and response probability function

Let ξ be a design in the design space \mathcal{X} containing all designs of interest. The design ξ is defined as a measure:

$$\xi = \left\{ \begin{array}{cccccc} x_1 & x_2 & \dots & x_j & \dots & x_q \\ n_1 & n_2 & \dots & n_j & \dots & n_q \end{array} \right\}, \quad \text{where } j = 1, \dots, q, \quad (4.2)$$

and the integers n_j indicate the number of subjects responding at time points x_j . The number of subjects responding to each x_j is usually equal, i.e., $n_1 = n_2 = \dots = n_j = \dots = n_q$ for complete data. However, if a dropout process arises, the number of subjects responding at x_j will monotonically decrease, i.e., $n_1 \geq n_2 \geq \dots \geq n_q$ where $n_1 > n_q$ and the probability of obtaining data at design point x_j will depend upon these design points. Let us denote a design with complete data by ξ_c and the design with possible dropout by ξ_d .

Let $p(x_j)$ be the response probability of measuring a subject at time points x_j . At the first design point x_1 of the study, the probability of

response is $p(x_1) = 1$ and the number of subjects responding at x_1 is equal to n_1 . At the j^{th} design point with response probability $p(x_j)$, the expected number of subjects $n_j = n_1 p(x_j)$. Finally, at the last design point x_q , the number of subjects n_q represents the number of subjects completing the experiment. For $j = q$, n_q is the number of subjects responding at all q design points. We will assume that at least one subject is observed at this last point, i.e., $n_q \geq 1$. It is assumed that $p(x_j)$ is a monotonically decreasing function with $p(x_1) \geq p(x_2) \geq \dots \geq p(x_q)$ where $p(x_1) > p(x_q)$.

It should be reminded that a response probability $p(x_j)$ at a design point x_j is complementary to a dropout process $1 - p(x_j)$ at the same design point x_j . An important assumption, though, is that the dropout occurs through a noninformative mechanism, i.e., responses are missing at random (MAR), see Little and Rubin (1987).

Different probability response functions $p(x_j)$ have been considered elsewhere (Imhof et.al., 2002, 2004; Galbraith and Marschner, 2002). In this study, however, we will focuss on two polynomial response functions, namely, a linear function $p_{lin}(x_j) = a_0 + a_1 x_j$ and a quadratic function $p_{quad}(x_j) = a_0 + a_1 x_j + a_2 x_j^2$ (where $0 \leq p(x_j) \leq 1$).

4.3.2. Linear mixed models with dropout

Let us discriminate between n_j subjects responding at time points x_j and m_j subjects having a total of j responses. This is done to account adequately for the contribution of m_j subjects having a similar response structure.

We introduce the response probability $p(x_j)$ and corresponding design matrices $X^{[j]}$ for the fixed-effects and $Z^{[j]}$ for the random-effects, as follows:

$X^{[q]}, Z^{[q]}$	are the design matrices of m_q subjects having complete measurements at all q design points: $p(x_q) = \frac{n_q}{n_1}$
$X^{[q-1]}, Z^{[q-1]}$	are the design matrices of m_{q-1} subjects having $(q - 1)$ measurements at design points x_1 to x_{q-1} : $p(x_{q-1}) = \frac{n_{q-1}}{n_1}$
\vdots	\vdots
$X^{[j]}, Z^{[j]}$	are the design matrices of m_j subjects having j measurements at design points x_1 to x_j : $p(x_j) = \frac{n_j}{n_1}$
\vdots	\vdots

$X^{[2]}, Z^{[2]}$ are the design matrices of m_2 subjects
 having measurements at x_1 to x_2 : $p(x_2) = \frac{n_2}{n_1}$
 $X^{[1]}, Z^{[1]}$ are the design matrices of m_1 subjects
 having only 1 measurement at x_1 : $p(x_1) = 1$.

In general, if the number of subjects having all responses to all q design points is denoted by $m_q = n_q$ and the number of subjects with only $j < q$ responses is denoted by the difference $m_j = n_j - n_{j+1}$, then the number of subjects m_j having j responses is given as:

$$m_j = \begin{cases} n_q & \text{if } j = q, \\ n_j - n_{j+1} & \text{if } j < q, \end{cases} \quad (4.3)$$

where m_j can also be written in terms of the response probability function $p(x_j)$, i.e., $m_q = n_q = n_1 p(x_q)$ for $j = q$ and $m_j = n_j - n_{j+1} = n_1 [p(x_j) - p(x_{j+1})]$ for $j < q$.

The design matrices $X^{[j]}$ and $Z^{[j]}$ ($j = 1, \dots, q$) are matrices of size $j \times p$ and $j \times k$, respectively, whose corresponding response covariance matrix reads as:

$$V^{[j]} = Z^{[j]} D Z'^{[j]} + \sigma^2 \Psi^{[j]}. \quad (4.4)$$

Because m_j subjects share identical $X^{[j]}$ and $Z^{[j]}$, the asymptotic covariance matrix of the best linear unbiased estimator $\hat{\beta}$ can be estimated by:

$$\text{Var}(\hat{\beta}) = \left(\sum_{j=1}^q m_j X'^{[j]} V^{[j]-1} X^{[j]} \right)^{-1}, \quad (4.5)$$

where m_j is defined as in (4.3) for all j .

The asymptotic covariance matrix (4.5) is equal to the inverse of the Fisher information matrix. As usual, small valued asymptotic covariance matrices of the parameter estimators correspond to large valued information matrices. The amount of information contained in a specific design and its precision is represented by this Fisher information matrix. Estimators with high precision have small valued estimated (asymptotic) variances.

4.3.3. D -Optimal designs

In this paper, we will focus on the determinant or D -optimality criterion. This criterion is very frequently used because of its scale invariance property

(Ouwens et.al., 2006), i.e. the D -optimal design for example in the interval $[0, +2]$ can be obtained by a transformation of the D -optimal design in the interval $[-1, +1]$. It also has a natural interpretation in terms of the volume of the simultaneous confidence ellipsoid of the fixed parameters (Atkinson and Donev, 1996).

Consider a set of W linear mixed models M_ω of the form (4.1), where $\omega = 1, \dots, W$. For a specific model M_ω , the determinant of the asymptotic variance covariance matrix or generalized variance of the estimated parameters $\hat{\beta}$ is:

$$\text{Det}[\text{Var}_{M_\omega}(\hat{\beta}|\xi)] = \text{Det} \left[\sum_{j=1}^q m_j X_\xi^{[j]} V_\xi^{[j]-1} X_\xi^{[j]} \right]^{-1}, \quad (4.6)$$

where ξ refers to the design that is used for the estimation of $\text{Var}(\hat{\beta})$ under model M_ω . Note that for the complete data ξ is symbolized here by ξ_c and for data with dropout by ξ_d .

The D -optimality criterion minimizes the generalized variance of the estimated parameters $\hat{\beta}$, see Atkinson and Donev (1996). Design ξ^* is called the D -optimal design for model M_w if it satisfies the inequality:

$$\text{Det}[\text{Var}_{M_w}(\hat{\beta}|\xi^*)] \leq \text{Det}[\text{Var}_{M_w}(\hat{\beta}|\xi)] \quad \forall \xi \in \mathcal{X}. \quad (4.7)$$

Notice that by minimizing this determinant or generalized variance, we actually maximize the power of the simultaneous test, i.e. all parameters in the vector β are zero.

4.3.4. Relative efficiency

To study the effect of dropout on the efficiency of D -optimal designs, two types of D -optimal designs are computed, i.e., ξ_c^* and ξ_d^* , respectively.

The design ξ_c^* is the D -optimal design for *complete data*: $\xi_c^* = \{x_1^{*c} \ x_2^{*c} \ \dots \ x_j^{*c} \ \dots \ x_q^{*c}; n_1 \ n_2 \ \dots \ n_j \ \dots \ n_q\}$, where x_j^{*c} are the D -optimal time points for ξ_c^* . Such a design assumes that the response probability $p(x_j^{*c}) = \frac{n_j}{n_1} = 1$, for all j . However, this design will be suboptimal if, in fact, there is dropout, i.e., if $p(x_j^{*c}) < 1$ for all j .

Let the design ξ_d^* be the D -optimal design for *data with dropout*. The design ξ_d^* has the form $\xi_d^* = \{x_1^{*d} \ x_2^{*d} \ \dots \ x_j^{*d} \ \dots \ x_q^{*d}; n_1 \ n_2 \ \dots \ n_j \ \dots \ n_q\}$. The time points x_j^{*d} are D -optimal for a dropout process $1 - p(x_j^d)$. The response probability is given by $p(x_j^{*d}) = \frac{n_j}{n_1}$ and $n_1 > n_2 > \dots > n_j >$

$\dots > n_q$.

Since our goal is to study the effect of an incorrect specification of the dropout process on the relative efficiency of D -optimal designs, it is assumed that the underlying model is also known, i.e., both optimal designs have the same variance-covariance structure, the same number of time points q and the same number of parameters p .

Loss of efficiency can be estimated by comparing the efficiency of a design ξ_c^* that is optimal for a model with complete data to that of an optimal design ξ_d^* for an underlying model with dropout. In other words, the generalized variance of the estimated regression coefficients $\text{Var}_{M_w}(\hat{\beta}|\xi_d^*)$ for the design ξ_d^* is compared to the $\text{Var}_{M_w}(\hat{\beta}|\xi_c^*)$ for the design ξ_c^* .

The loss of efficiency of the optimal design ξ_c^* when instead an optimal design ξ_d^* should have been chosen can thus be calculated by:

$$RE(\xi_c^*|\xi_d^*) = \left[\frac{\text{Det}\{\text{Var}_{M_w}(\hat{\beta}|\xi_d^*)\}}{\text{Det}\{\text{Var}_{M_w}(\hat{\beta}|\xi_c^*)\}} \right]^{\frac{1}{p}} = \left\{ \frac{\text{Det}(\sum_{j=1}^q m_j^d X_{\xi_d^*}^{[j]} V_{\xi_d^*}^{[j]-1} X_{\xi_d^*}^{[j]})^{-1}}{\text{Det}(\sum_{j=1}^q m_j^c X_{\xi_c^*}^{[j]} V_{\xi_c^*}^{[j]-1} X_{\xi_c^*}^{[j]})^{-1}} \right\}^{\frac{1}{p}}. \quad (4.8)$$

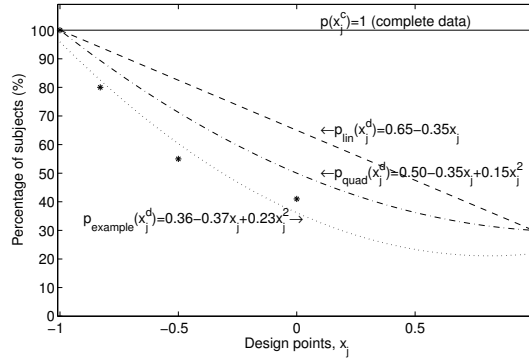
where $m_j^d = n_1[p(x_j^{*d}) - p(x_{j+1}^{*d})]$ and $m_j^c = n_1[p(x_j^{*c}) - p(x_{j+1}^{*c})]$ if $j < q$. Moreover, $m_j^c = n_q^c$ as $m_j^d = n_q^d$ happen to be all equal if $j = q$.

Notice that the D -optimal design ξ_d^* is optimal for the model with dropout and that, the numerator of (4.8) has a smaller variance than the denominator. The number of fixed parameters of the model is equal to p . Taking the ratio of the determinants in (4.8) to the $1/p$ power, results in an efficiency measure which is proportional to design size, irrespective of the dimension of the model (Atkinson and Donev, 1996; Pukelsheim, 1993). The interpretation of the relative efficiency is straightforward, e.g. $RE = \frac{1}{2}$ implies that two replicates of an optimal design ξ_c^* for the complete data case will be as efficient as one replicate of the optimal design ξ_d^* for the data with dropout. Relative efficiencies of 0.85 or higher will usually be preferred.

4.4. Numerical analysis

In this section the effect of dropout on the location of D -optimal design points will numerically be investigated. We will introduce two dropout functions and we will describe four classes of linear mixed models under consideration. Thereafter, the D -optimal designs for these four model classes and the two dropout processes will be presented along with the D -optimal

Figure 4.1: Response probability function, $p(x_j)$. Solid line: Constant $p(x_j^c)$. Dashed line: $p_{\text{linear}}(x_j^d)$. Dashed-dot Line: $p_{\text{quad}}(x_j^d)$. Star: Data example. Dotted line: Fit $p_{\text{example}}(x_j^d)$



designs for complete data. Finally, the efficiency loss of using D -optimal designs based on the incorrect assumption of complete data instead of those designs with dropout will be displayed by means of relative efficiency plots.

4.4.1. Dropout function

Four response probability functions are displayed in Figure 4.1. Notice that the response probability function $p(x_j)$ is always complementary to the dropout process, i.e., $1 - p(x_j)$.

In Figure 4.1, the x -axis represents the location of the design points x_j on the interval $[-1, 1]$ and the y -axis corresponds to the response probability in terms of percentage or the number of subjects in the trial and at the design point x_j . The solid line corresponds to the probability function $p(x_j^c) = 1$ of the complete data case, the dashed line represents the linear response probability $p_{\text{lin}}(x_j^d) = 0.65 - 0.35x_j$ and the dashed-dot line correspond to the quadratic response probability $p_{\text{quad}}(x_j^d) = 0.50 - 0.35x_j + 0.15x_j^2$. The stars represent the data on cocaine addicts of the randomized clinical trial data for section 4.2 and the dotted line is the corresponding polynomial fit.

4.4.2. Classes of linear mixed models

Linear mixed models in the following four classes were considered. For these classes we have restricted ourselves to $p = 2, 3$ and 4 fixed polynomial parameters with $q = 3$ and 4 time points. The autocorrelation of the error terms ranged from $\rho = 0$ to 0.80. In the four classes we consider random models with different values for ρ and d_{11}, d_{12} and d_{22} .

Class 1: Fixed effects models. For these models the variance-covariance matrix D of the random parameters is a null matrix.

Class 2: Random intercept models. The variance-covariance matrix D is now a scalar with values ranging between $0 < d_{11} \leq 10$.

Class 3: Random intercept and slope models. The variance-covariance matrix D is a 2×2 diagonal matrix with elements $0 < d_{11} \leq 10$ and $0 < d_{22} \leq 3$.

Class 4: Random intercept and slope models with covariances between the random effects equal to $d_{12} = \rho^* \sqrt{d_{11} \times d_{22}}$, where the correlation between the random intercept and slope is $\rho^* = 0.8$. Without loss of generality, we have considered only positive values for ρ^* (Ouwens et.al., 2002).

In these classes of linear mixed models, we found that, in general, the trends of the optimal allocation of the time points with dropout depended more on the form of the variance-covariance matrix D than on the specific values of its individual elements d_{11} , d_{22} and d_{12} . Therefore, we will only present the results for models with $p = 2, 3$ and 4 fixed polynomial parameters and $q = 3$ and 4 time points. The elements of D are selected depending on the model class under consideration, i.e., $d_{11} = \{0, 1\}$, $d_{22} = \{0, 3\}$ and $d_{12} = \{0, 0.8\sqrt{3}\}$.

4.4.3. Algorithm

The D -optimal designs for the complete and for the incomplete data case connected to the classes of linear models described in section 4.4.2 are numerically derived by optimizing the generalized variance of their estimated parameters. That is evaluating equation (4.6) subject to constraint (4.3). A program using Matlab's function FMINSEARCH was implemented in Matlab version 7.0.1(R14). The program uses a multidimensional unconstrained nonlinear minimization algorithm developed by Nelder and Mead (Lagarias et.al, 1998; Matlab, 2004). The program requires an initial design value ξ_0 . We used equally-spaced designs as initial values.

4.4.4. D -optimal designs for data with dropout

In Figures 4.2-4.5, the D -optimal designs for complete data and data with dropout are plotted. Each figure contains 12 subplots. The first-, second- and third-column of each figure display the D -optimal designs for complete data, data with linear dropout and data with quadratic dropout, respectively. The rows of each figure correspond to the form of the variance-covariance matrix D of the random parameters, i.e., *Class 1*, *2*, *3* and *4*, respectively. Moreover, in each subplot, the x -axis refers to the location

x_j of the design points in the interval $[-1, 1]$ and the y -axis represents the serial correlation parameter ρ within the subjects. As already indicated, the range of this parameter is $[0, 0.8]$.

D -optimal designs for the case of $q = 3$ and $q = 4$ time points fitted by a first-degree polynomial ($p = 2$) are shown in Figures 4.2 and 4.3, respectively.

In general, D -optimal designs for data with dropout that correspond to the *Class 1* and *Class 2* are moderately displaced as compared to the D -optimal designs for complete data. Notice that the optimal designs for complete data are similar to those obtained by Tan and Berger (1999). Moreover, the direction of the displacements seems to depend on the sort of dropout involved. The D -optimal designs connected to a linear response probability function $p_{\text{lin}}(x_j^d)$ tend to move rightwards and those optimal designs corresponding to a quadratic function $p_{\text{quad}}(x_j^d)$ tend to move leftwards. We also observed that D -optimal designs for data with dropout corresponding to the *Class 3* and *Class 4* models are more severely displaced. Strikingly, the direction of these displacements do not depend on the sort of dropout involved. Almost all D -optimal designs of the *Class 3* and *Class 4* models move leftwards with respect to the D -optimal designs for complete data.

The optimal location of the cases of $q = 3$ and $q = 4$ time points fitted by a second-degree polynomial ($p = 3$) are displayed in Figures 4.4 and 4.5, respectively. It can be observed that, regardless the variance-covariance matrix of random parameters involved, all D -optimal designs for data with dropout move leftwards of the optimal designs for complete data.

In the case of D -optimal designs for $q = 4$ time points fitted by a third-degree polynomial ($p = 4$) (plots are not shown here), it was observed that there are no significant differences on the location of D -optimal designs. The D -optimal designs for data with linear and quadratic dropout resemble very much those designs for complete data. Thus, nor the model class or the autocorrelation parameter seem to affect the location of the encountered D -optimal designs for data with dropout.

4.4.5. Efficiency loss due to the incorrect assumption of complete data

In Figure 4.6, the relative efficiencies (RE) of D -optimal designs with linear and quadratic response probability functions with respect to D -optimal designs for complete data are displayed. This figure contains 6 subplots. The subplots of the two columns display the relative efficiencies $RE(\xi_c^*|\xi_d^*)$ whenever an incorrect D -optimal design for complete data is assumed instead

Figure 4.2: D -optimal designs for $p = 2, q = 3$

$$p(x_j^c) = 1 \quad p_{\text{lin}}(x_j^d) = 0.65 - 0.35x_j \quad p_{\text{quad}}(x_j^d) = 0.50 - 0.35x_j + 0.15x_j^2$$

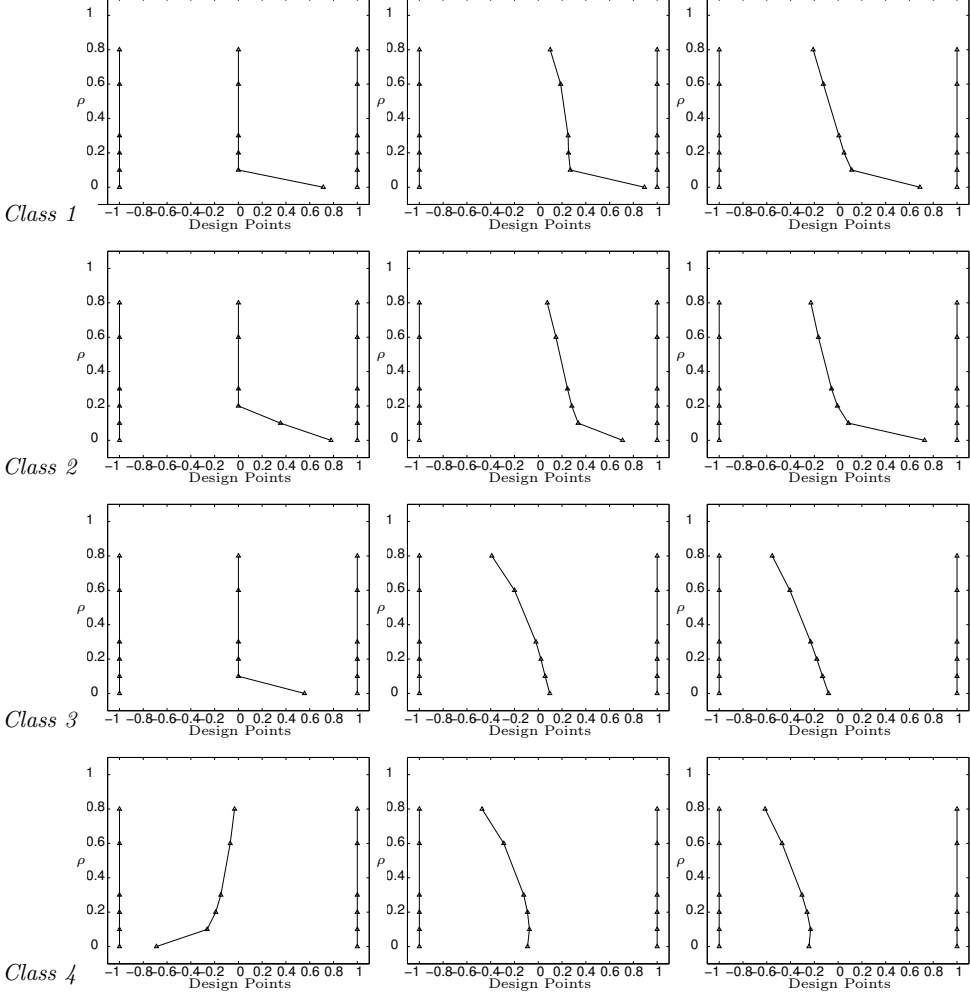


Figure 4.3: D -optimal designs for $p = 2, q = 4$

$$p(x_j^c) = 1 \quad p_{\text{lin}}(x_j^d) = 0.65 - 0.35x_j \quad p_{\text{quad}}(x_j^d) = 0.50 - 0.35x_j + 0.15x_j^2$$

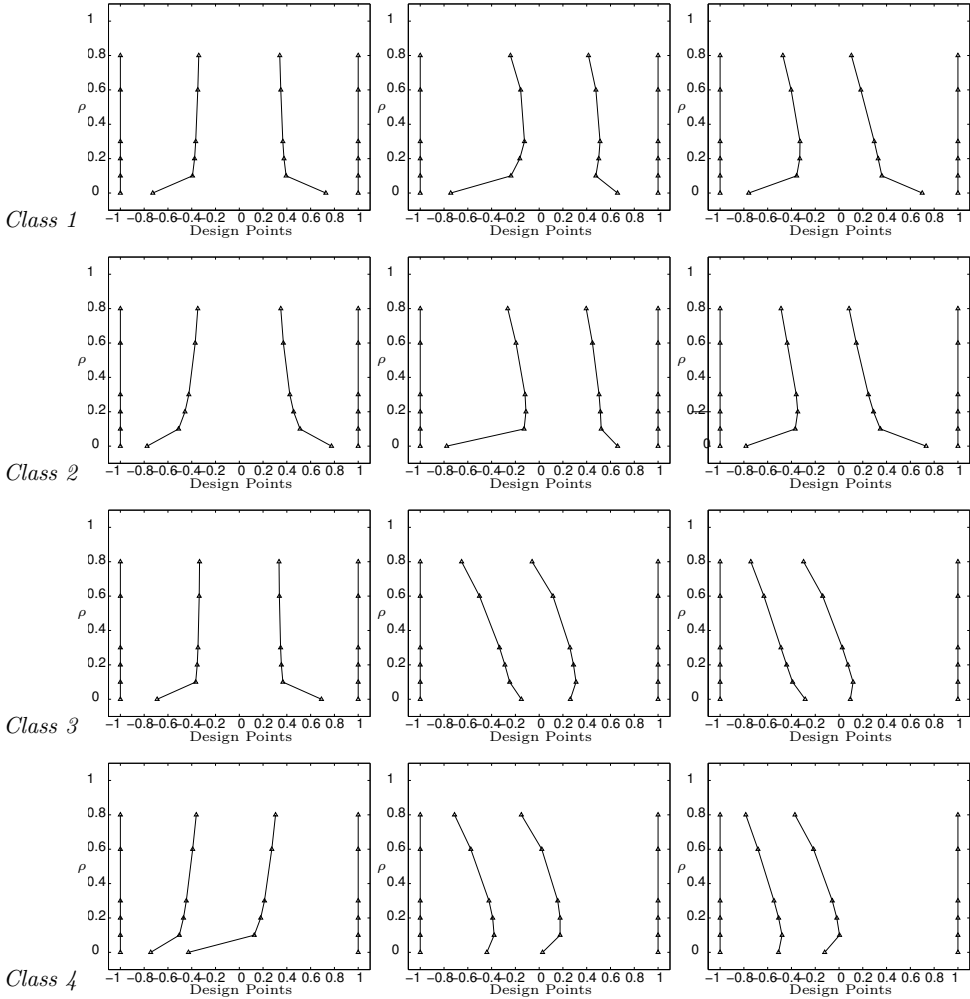


Figure 4.4: D -optimal designs for $p = 3, q = 3$

$$p(x_j^c) = 1 \quad p_{\text{lin}}(x_j^d) = 0.65 - 0.35x_j \quad p_{\text{quad}}(x_j^d) = 0.50 - 0.35x_j + 0.15x_j^2$$

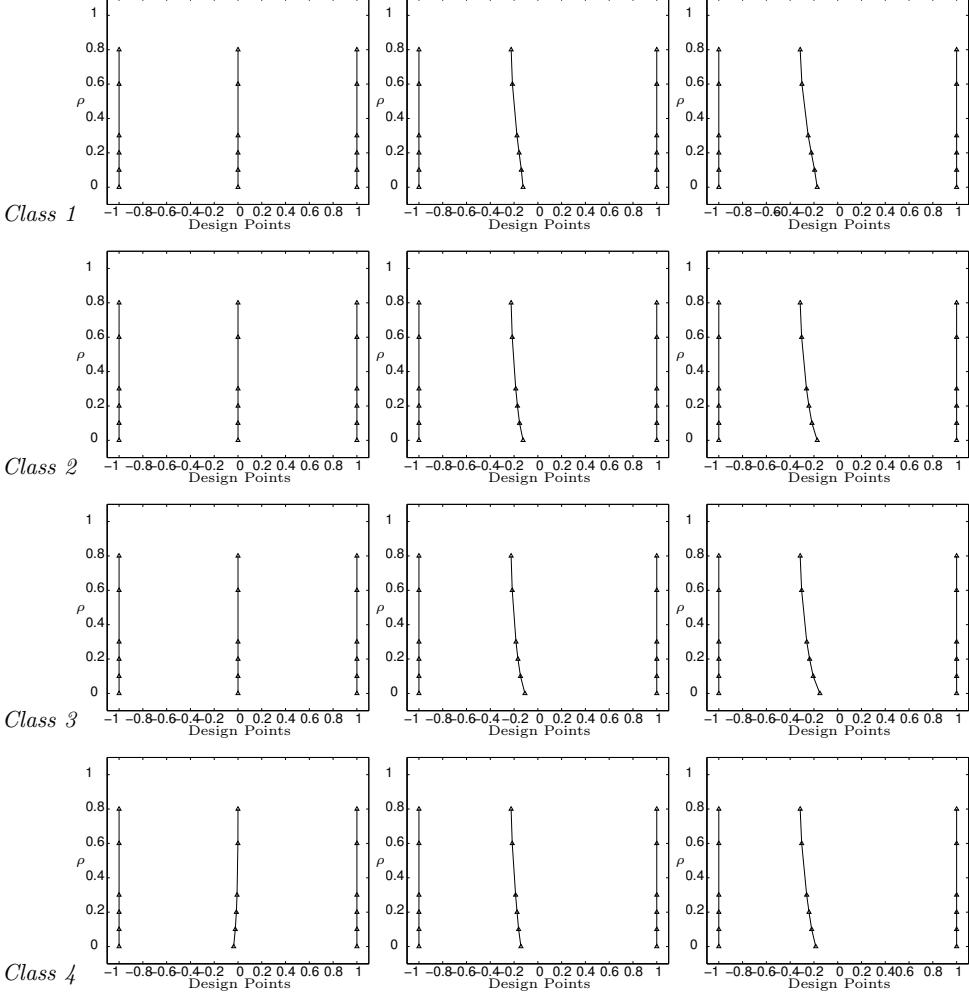
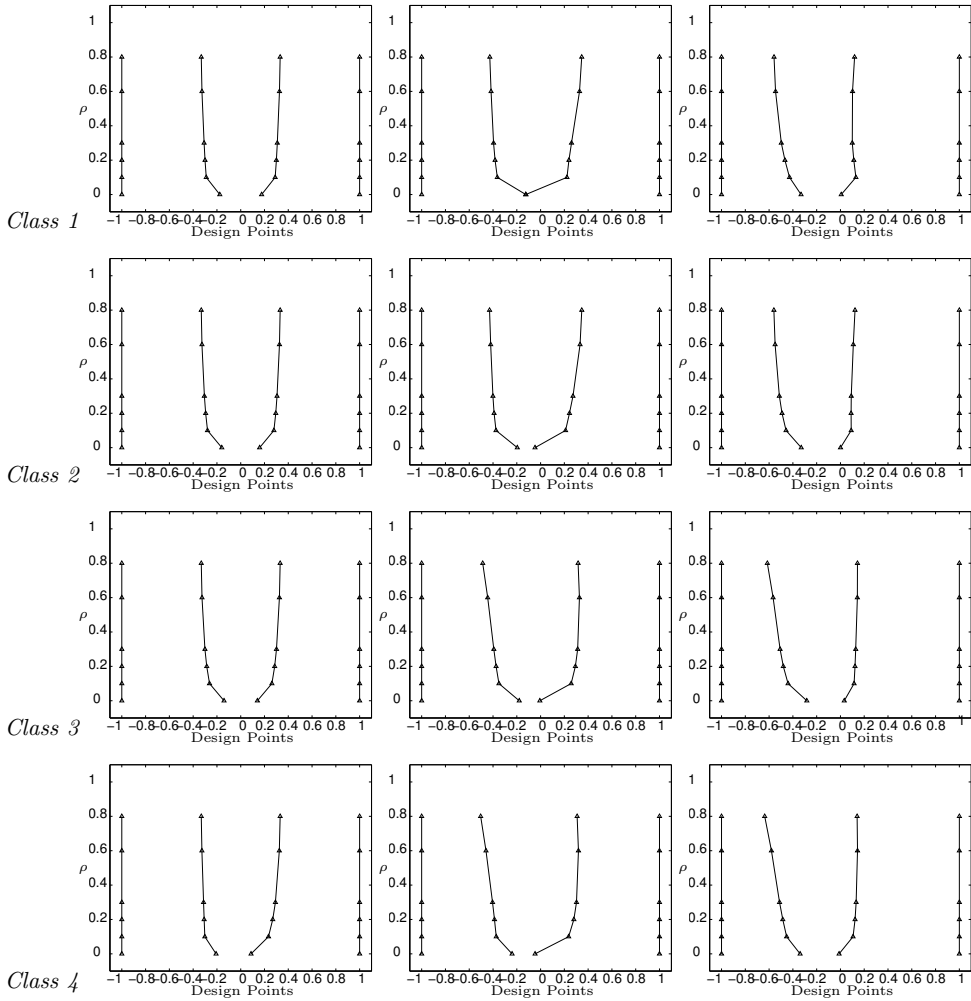


Figure 4.5: D -optimal designs for $p = 3, q = 4$

$$p(x_j^c) = 1 \quad p_{\text{lin}}(x_j^d) = 0.65 - 0.35x_j \quad p_{\text{quad}}(x_j^d) = 0.50 - 0.35x_j + 0.15x_j^2$$



of those D -optimal designs for data with a linear and quadratic dropout process, respectively.

The RE's for the first-degree polynomial ($p = 2$) for the case of $q = 3$ and 4 time points are plotted in the first- and second-row, respectively. In the same figure, the RE's for the second-degree polynomial ($p = 3$) for the case of $q = 3$ and 4 time points are displayed in the third and fourth rows, respectively. We have not plotted the efficiency plots for the third degree-polynomial ($p = 4, q = 4$) because its efficiency loss is negligible.

In each subplot, the x -axis represents the models M_w , where $w = 1, \dots, 24$. These models are selected as illustration of each class according to the variance-covariance matrix D of the random parameters, i.e., based on *Class 1, 2, 3* and *4*. The y -axis represents the relative efficiency. Moreover, an horizontal line at $RE = 0.85$ is drawn as reference line for comparisons's sake.

It is interesting to notice that both columns of subplots are practically alike. This shows that, the RE does not depend on the involved dropout function. By looking at the first- and second-row of subplots, it is seen that *Class 3* and *Class 4* models having a first-degree polynomial form display the largest loss of efficiency. However, all relative efficiencies remain equal or higher than 0.85. We may conclude that for these models, data with dropout as large as 70% will not lead to efficiency losses higher than 15%. Finally, notice that the relative efficiencies for the second-degree polynomial for $q = 3$ and 4 time points are very high.

4.4.6. D -optimal designs and efficiency loss of the randomized clinical trial example

The follow-up study on cocaine addicts reported by Carroll et.al. (1994) and Nich et.al. (1997) was based on five points ($q = 5$) and was fitted by a first-degree polynomial ($p = 2$) with the assumption of a first-order autoregressive error covariance structure. These authors fitted a linear mixed model of *Class 4* to the data. The estimated variance-covariance matrix D of the random parameters was $D = (0.030 \quad -0.004; -0.004 \quad 0.005)$. The estimated correlation between random intercept and slope variances was $\rho^* = -0.328$. Time is transformed by a logistic transformation $\log(\text{month}+1)$ to account for unequally spaced intervals between follow-up evaluations (0, 1, 3, 6, 12 months), see Carroll et.al. (1994, p.994) and Nich et.al. (1997, p.256). Thus, their design (after transformation and re-scaling) was $\xi_{\text{example}} = (-1 - 0.4596 \quad 0.0810 \quad 0.5173 \quad 1)$.

The location of the D -optimal designs points for $q = 5$ time points

Figure 4.6: Relative Efficiencies for $p = 2, 3$ and $q = 3, 4$ time points between the dropout case with linear probability function $p(x_j^d) = 0.65 - 0.35x_j$ and the complete data case; and, the quadratic probability function $p(x_j^d) = 0.50 - 0.35x_j + 0.15x_j^2$ and the complete data case.

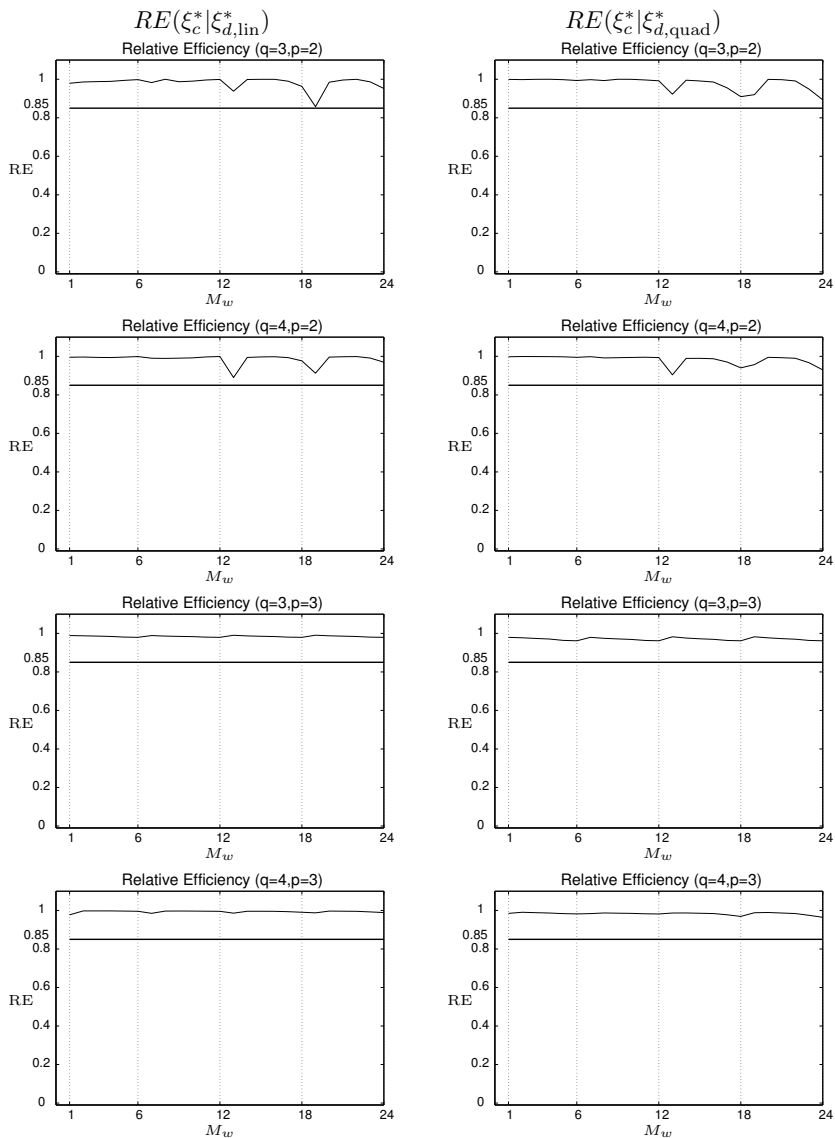
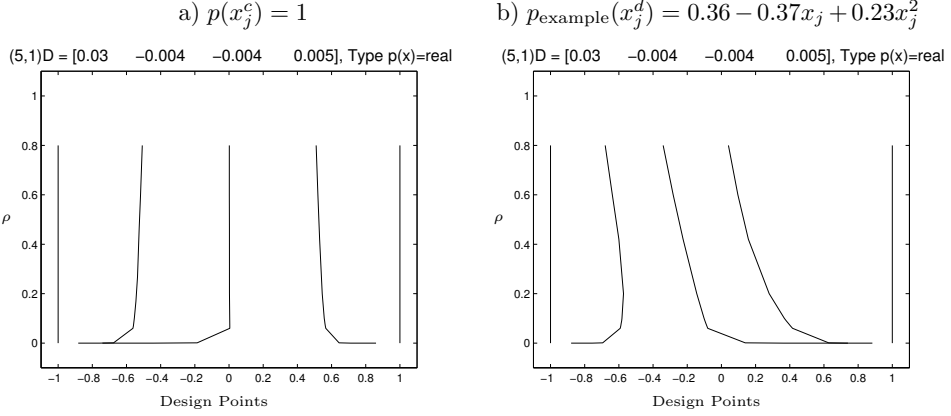


Figure 4.7: D -optimal designs of *Class 4* for $p = 2$, $q = 5$.

fitted by a first-degree polynomial ($p = 2$) is shown in Figure 4.7. The left and right subplots correspond to optimal designs for complete data and for data with dropout process as shown in Figure 4.1.

It can be seen that optimal design points for data with dropout are moderately displaced with respect to those for complete data and that the direction of the displacements is leftwards. Moreover, the relative efficiencies of choosing the D -optimal design $\xi_{c,\text{example}}^*$ while instead the design $\xi_{d,\text{example}}^*$ should have been used are in the range $[0.9965, 0.9788]$ for ρ in the range $[0, 0.80]$. This loss of efficiency is very small.

However, if we compute the relative efficiencies of choosing the design ξ_{example} while instead its D -optimal design $\xi_{d,\text{example}}^*$ should be selected, the $RE(\xi_{\text{example}}|\xi_{d,\text{example}}^*)$ drops to the range $[0.8901, 0.8245]$ for ρ in the range $[0, 0.80]$. Notice that the RE of the correlated cases are somewhat lower than the reference value of 0.85.

4.5. Conclusions

In this paper, we have focused on how does the efficiency of optimal designs with dropout compare with those without dropout. We show that D -optimal designs are robust against dropout and that these designs can render improvements in longitudinal studies which struggle with time-consuming or unstable processes and where minimizing the number of sessions and the number of subjects involved are crucial.

The effect of dropout on the efficiency of D -optimal designs of linear mixed models can be summarized as follows:

The displacement of the D -optimal time points due to dropout is largest

for linear mixed models having a first-degree polynomial ($p = 2$) and $q = 3$ and $q = 4$ time points. The displacement of time points becomes smaller as the degree of the polynomial increases to $p = 3$ and $p = 4$. The loss of efficiency when a D -optimal design ξ_c^* is incorrectly used instead of the D -optimal design ξ_d^* for data with expected dropout, is also greatest for a linear ($p = 2$) polynomials than for $p = 3$ and $p = 4$ degree polynomials. It was shown that, the lowest RE 's correspond to *Class 3* and *4* models with $\rho = 0$. However, the relative efficiency remains very high for all the linear mixed effects models that were studied.

These results were in coherence with the efficiency loss computed from the data of the clinical trial example. Maximum efficiency loss is found to be smaller than 3% ($RE = 0.9788$). The computations also showed that the original design used by Carroll et.al. (1994) and Nich et.al. (1997) has a $RE = 0.83$ when compared with the D -optimal design corresponding to the response probability function for their data.

These results seem to support the conclusion that for the linear mixed models considered in this paper dropout does not seem to be a serious problem in the design stage of a study. If one uses a D -optimal design for complete data, while a dropout is encountered in the data collection stages, efficiency loss does not exceed 15% and researchers can in these cases easily compensate for such small efficiency loss by increasing the initial sample by at most 15%.

Notice that the variation in the response probability function seems to be neglected, i.e. the observed proportion of individuals at each observation point is exactly equal to the assumed dropout probability function. However, we have tried different dropout functions and the results obtained showed similar small losses of efficiency. These results are based on linear and quadratic response probability functions however, other functions could instead have been employed (Herzberg and Andrews, 1976a; Hackl, 1995; Verbeke and Molenberghs, 2000).

Our results are limited in the sense that we assumed probability response functions indicating a dropout of at most 70% and our computations were based on the assumption that the end points of the study remained fixed. It seems worthwhile to extend these results to cases where the design space is not fixed between $[-1, +1]$. Incorporation of 100% dropout at the later time points would lead to relaxation of the fixed endpoint condition.

Moreover, since researchers could be interested in a particular parameter or a subset of parameters, other criteria like D_s - or c -optimality could also be of interest (Atkinson and Donev, 1996, ch.21). In practice, the timing of measurements is also unlikely to be left to purely statistical considerations.

In many cases non-statistical considerations, i.e. logistical complexity, trial safety, dose-response considerations, are likely to play an important role in deciding on a certain design.

Finally, the results of this paper are based on models with first-order auto-regressive covariance structures. Since any dropout process could change the assumed correlation structure of the observations between the different time points, it is important to extend these results for other correlations structures like compound symmetry or Toeplitz forms. Further research in these directions is strongly encouraged and the general approach shown in this article could be applied as well.

CHAPTER 5

On the scale dependency of A –optimal designs for random intercept models with serially correlated errors ¹

¹*Submitted for publication*

Abstract

In this paper, the robustness of A -optimal designs for serially correlated random intercept models is numerically investigated with respect to their design scale and connected weights. The relative efficiency is used to measure the loss of efficiency when designs that are optimal in one scale are applied to another scale. Small to moderate efficiency losses arise when designs which are optimal in the narrow scales: $[-1, 1]$, $[0, 1]$ and $[0, 2]$, are interchanged. Moderate to large efficiency losses are reported when designs which are optimal in a narrow scale are applied to a broad scale $[0, 10]$ and vice-versa. Moreover, the performance of equally-weighted A -optimal designs is not good when compared to their connected designs with unequal weights. Designs which are optimal in the scale $[0, 10]$ are the least efficient.

Keywords: A -optimality; First-order autocorrelation; Random-intercept model; Relative efficiency; Robust designs

5.1. Introduction

In health and medical sciences longitudinal designs are often used to measure a particular characteristic over time. A commonly used model to analyse such time-structured data is the well-known linear mixed-effects model (Verbeke and Molenberghs, 2000). Methods to choose a proper (optimal) design to collect time structured data have been studied extensively (Abt et.al., 1997, 1998; Atkins and Cheng, 1999; Bischoff, 1993; Cheng, 1999; Mentre et al., 2006; Tan and Berger, 1999; Ouwens et.al., 2002). These methods are usually based on the so-called D -optimality criterion. One of the reasons for using this D -optimality criterion is its invariance with respect to the scale of the time points.

A second criterion that is often used to evaluate optimal designs is the A -criterion. This criterion is perhaps more appealing to practitioners because it is based on the idea of making the average variance of the parameter estimators of the model as small as possible. This criterion has been studied in the literature for fixed-effects models with independent errors (Elving, 1959; Fedorov, 1972; Kiefer, 1975; Colinsk and Watts, 1979; Pukelsheim, 1993; Wong, 1994; Dette, 1997). A -optimal designs for fixed-effect models for various polynomial degrees in the design interval $[-1, 1]$ have been published by Pukelsheim (1993).

Recent efforts on fixed-effects models with correlated errors have been carried out by Abt et.al. (1998) and Moerbeek (2004). Abt et.al. described the weights of A -optimal designs for the intraclass and the serial correlation structures in the interval $[-k, k]$ and Moerbeek reported the location of A -optimal design points obtained for models with serially correlated errors under the restriction of equal weights in the interval $[0, 2]$.

An important disadvantage of the A -optimal designs is that they are scale-dependent, i.e. they depend on their underlying scale. In longitudinal studies, for example, if a linear transformation is considered to change the time scale of a research study from the time interval $[-1, 1]$ into $[0, 10]$, the A -optimal designs for the interval $[0, 10]$ cannot be obtained by applying the same linear transformation to the A -optimal designs for the time interval $[-1, 1]$. Moreover, tables for A -optimal designs covering a sufficiently large range of interval parameter values a, b are not feasible. This drawback becomes particularly important when time-structure data are encountered, since the time scale may be quite different for different studies and because of the large variety of models used in health sciences and medicine.

The problem of scale-dependency of A -optimal designs for fixed-effects models has been addressed in the literature (Fedorov, 1972; Kiefer, 1975;

Pukelsheim, 1993; Dette, 1997). Dette (1997) proposed the use of standardized criteria and provided some results for standardized A -optimal designs in polynomial regression.

A second issue in using the A -optimality criterion is that A -optimal designs for fixed-effects models with independent errors are known to be unequally weighted. See Pukelsheim (1993), among others. Although, the restriction of equal weights is usually made in repeated measurement studies (Moerbeek, 2004), it is not known whether such designs are highly efficient, compared to the optimal designs with unequal weights.

The purpose of this study is three-fold. First A -optimal designs for the random-intercept model will be evaluated for different parameter values. Then the A -optimal designs will be compared to each other for different time scales to investigate whether these designs are in some way robust with respect to different time scalings. Finally, efficiency loss will be investigated when A -optimal designs are restricted to equal weighted designs.

The rest of the paper is organized in the following way. In section 5.2, the random-intercept model is presented along with the A -optimality criterion and the relative efficiency. In section 5.3, two numerical approaches to compute A -optimal designs are introduced. In this section the choices made with regard to various combinations of design intervals and model parameters are also explained. In section 5.4, the obtained A -optimal designs are presented along with relative efficiency plots. In the concluding section 5.5, our results are compared with those found in the literature and a discussion follows.

5.2. Random-intercept model with serially correlated errors

Let the vector $\mathbf{y}_i = (y_{i1}, \dots, y_{iq_i})^T$ represent the repeated measurements of the response of subject i^{th} , where the number of repeated measurements on subjects are not necessarily the same, i.e. $j = 1, \dots, q_i$. The linear model with random-effects for subject i^{th} is given by

$$\mathbf{y}_i = X_i \boldsymbol{\beta} + \mathbf{1}_i b_{0i} + \mathbf{e}_i, \quad (5.1)$$

where $\boldsymbol{\beta}$ is a $(p \times 1)$ vector of population-average regression parameters called fixed-effects and where b_{0i} is a scalar representing the subject-specific intercept parameter often addressed to as a random-effect. The parameter b_{0i} represents how the intercept of the i^{th} subject deviates from the average population. This parameter is normally distributed with mean 0 and

random-intercept variance d_{11} . The matrices X_i and $\mathbf{1}_i$ are the $(q_i \times p)$ and $(q_i \times 1)$ design matrices for the fixed parameters β and the random parameters b_{0i} , respectively.

The vector \mathbf{e}_i represents the within-subjects error. In longitudinal data, this term is often serially correlated. The errors \mathbf{e}_i are assumed to be normally distributed as $N(0, \sigma^2 \Psi_i)$, where Ψ_i is a positive-definite covariance matrix which depends on the correlation parameter ρ of the time lag between the j^{th} and j'^{th} time points, i.e. x_{ij} and $x_{ij'}$, respectively. This structure is referred to as first-order autoregressive, i.e. AR(1). The elements of Ψ_i are $\rho^{\text{abs}(x_{ij} - x_{ij'})}$, where design points x_{ij} and $x_{ij'}$ are not necessarily equally spaced.

The best linear unbiased estimator of β has an asymptotic variance-covariance matrix

$$\text{Var}(\hat{\beta}) = \left(\sum_{i=1}^n X_i^T V_i^{-1} X_i \right)^{-1}, \quad (5.2)$$

where V_i is the variance-covariance matrix of the repeated measurements \mathbf{y}_i . The variance-covariance matrix V_i for the i^{th} subject having q_i repeated measurements in the design interval $[-1, 1]$ has the form

$$V_i = \begin{bmatrix} d_{11} + \sigma^2 & \dots & d_{11} + \sigma^2 \rho^{|-1-x_{ij}|} & \dots & d_{11} + \sigma^2 \rho^2 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ d_{11} + \sigma^2 \rho^{|x_{ij}+1|} & \dots & d_{11} + \sigma^2 & \dots & d_{11} + \sigma^2 \rho^{|x_{ij}-1|} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{11} + \sigma^2 \rho^2 & \dots & d_{11} + \sigma^2 \rho^{|1-x_{ij}|} & \dots & d_{11} + \sigma^2 \end{bmatrix}. \quad (5.3)$$

5.2.1. Design, A-optimality criterion and relative efficiency

The design ξ is a measure that represents the location of q design points and the weights w_j connected to each design point x_j , where $j = 1, \dots, q$. The parameter q represents the maximum number of design points observed among all n subjects. i.e. $q = \max_{i=1, \dots, n} q_i$. This implies that there is at least one person with all q measurements. The design ξ is an element of the design space Ξ and is defined as:

$$\xi = \left\{ \begin{array}{cccccc} x_1 & x_2 & \dots & x_j & \dots & x_q \\ w_1 & w_2 & \dots & w_j & \dots & w_q \end{array} \right\}, \quad \text{where } j = 1, \dots, q. \quad (5.4)$$

The weights w_j represent the relative proportion of the total observations taken at the x_j time points, and are subject to the constraint $\sum_{j=1}^q w_j = 1$. All x_j 's are located in the design interval $[a, b]$.

In this paper the A -optimality criterion will be investigated. This criterion is based on the main diagonal elements of $\text{Var}(\hat{\beta})$ in equation (5.2) i.e. on the variances of the estimators of the parameters in β . An A -optimal design ξ^* is the design for which the trace of $\text{Var}(\hat{\beta})$ is minimized among that of all possible designs ξ in the design space Ξ , i.e.

$$\text{Trace}\{\text{Var}(\hat{\beta}|\xi^*)\} \leq \text{Trace}\{\text{Var}(\hat{\beta}|\xi)\}, \quad \text{for all } \xi \in \Xi. \quad (5.5)$$

The efficiency of any design ξ can be compared to the efficiency of the A -optimal design ξ^* by the relative efficiency measure,

$$RE(\xi|\xi^*) = \left[\frac{\text{Trace}\{\text{Var}(\hat{\beta}|\xi^*)\}}{\text{Trace}\{\text{Var}(\hat{\beta}|\xi)\}} \right]. \quad (5.6)$$

The interpretation of the relative efficiency (5.6) is straightforward. For example, two replicates of a design ξ for which $RE = 0.5$ would be as efficient as one replicate of the A -optimal design ξ^* .

In this paper, we will use the relative efficiency (5.6) to compare designs in two different contexts and to answer two questions.

In the first context, the following question is answered: how much efficiency is lost when a design ξ_1^* , which is A -optimal in a first interval $[a, b]$ is applied to a second interval $[c, d]$, where the A -optimal design ξ_2^* holds? To measure how much efficiency is lost when design ξ_1^* is used in the interval $[c, d]$, we compute the ratio (5.6) between the optimal trace of $\text{Var}(\hat{\beta}|\xi_2^*)$ and the sub-optimal trace of $\text{Var}(\hat{\beta}|\xi_1^*)$.

The second context is concerned with the question: how much efficiency is lost if the A -optimal design under the restriction of equal weights ξ_{eq}^* is applied, while the A -optimal design with unequal weights ξ_{un}^* should be used? In this case, the ratio (5.6) is computed between the optimal trace $\text{Var}(\hat{\beta}|\xi_{\text{un}}^*)$ and the sub-optimal trace $\text{Var}(\hat{\beta}|\xi_{\text{eq}}^*)$. Notice that in this context both designs are optimal in the same design interval $[a, b]$ but ξ_{eq}^* is obtained under the additional constraint that all w_j s are equal.

5.3. Numerical method

Since A -optimal designs for random intercept models are difficult to derive analytically (Abt et.al., 1998), we will evaluate them by two numerical procedures.

Optimization Algorithms

In the first procedure, all possible combinations of weights w_j ($j = 1, \dots, q$) are systematically generated. Due to this systematic generation, the designs may have equal or unequal weights. Next, A -optimal designs ξ^* are obtained by evaluating stepwise the equation (5.5). This procedure is speeded up by allowing an initial optimization that relies first on a rough mesh that spans the whole design interval $[a, b]$. Thereafter, a second finer mesh that spans a more narrow interval $[a', b']$ around the *rough* optimal design points x_2^* is executed.

In the second procedure, optimal designs ξ_{eq}^* are obtained with the restriction of equal weights, $w_j = w$ ($j = 1, \dots, q$). The optimal location of time points x_j^* is found by evaluating the equation (5.5) too. The chosen nonlinear minimization algorithm was developed by Nelder and Mead and is implemented in MATLAB version 7.0.1(R14). For further details see e.g. Matlab (2004) and Molenberghs et al. (2006). Notice that the first procedure is more computationally intensive than the second. This method was chosen because it has a high convergence rate provided that the starting vector is in the neighborhood of the optimum. Different starting points are used to avoid local optimality.

Input Parameters

In this paper, the A -optimal designs for first- ($p = 2$) and second- ($p = 3$) order polynomial models with correlated errors are reported. In our numerical results, the input values of the model parameters are the random intercept variance $d_{11} = 0, 1, 2, 3, 4, 10$ and the autocorrelation parameter ρ in the range $[0, 0.8]$. Without loss of generality, the error variance parameter σ^2 was set at 1.

Choice of design intervals

Several different intervals for the design points have been used in the literature. Kiefer (1975) and Pukelsheim (1993), for example, have used the interval $[-1, 1]$; Dette and O'Brien (1999) have used $[0, 1]$ and Moerbeek (2004) has considered the interval $[0, 2]$. In this paper, we will use all these intervals along with the interval $[0, 10]$ which is used to show the specific effect of a large-range interval.

5.4. Results

A-optimal designs ξ_{un}^* and ξ_{eq}^* are reported for the design intervals $[-1, 1]$, $[0, 1]$, $[0, 2]$ and $[0, 10]$. As indicated before, A-optimal designs depend on the chosen design interval, i.e., an A-optimal design for the interval $[0, 1]$ cannot be linearly transformed to the A-optimal design in the interval $[0, 10]$. In the section *Numerical approximation under equal weights restriction*, it will be considered how, under certain conditions, a linear transformation for obtaining ξ_{eq}^* in different scales could be estimated.

5.4.1. A-Optimal designs

Algebraic solution for the first-order polynomial model with uncorrelated errors

For the case of linear regression with uncorrelated errors, the optimization problem of minimizing the $\text{Trace}\{\text{Var}(\hat{\beta})\}$ has a simple form. See Dette (1997).

Let x_j be a continuous variable in the interval $[a, b]$ where $b > a$. The trace of the variance-covariance of the parameter estimates is given by

$$\text{Trace}\{\text{Var}(\hat{\beta})\} = \frac{b^2 + 1}{(a - b)^2} \frac{1}{w_1} + \frac{a^2 + 1}{(a - b)^2} \frac{1}{w_2} \quad (5.7)$$

$$w_1 + w_2 = 1 \quad (5.8)$$

The optimal weights w_1^* and w_2^* at the endpoints of the interval are optimal at $x_1^* = a$ and $x_2^* = b$. The optimal weight w_1^* is

$$w_1^* = \frac{\sqrt{1 + b^2}}{\sqrt{1 + a^2} + \sqrt{1 + b^2}}, \quad (5.9)$$

and $w_2^* = 1 - w_1^*$. Equation (5.9) is equivalent to that reported by Dette (1997, eq. 4.2, p.106), but for illustration purposes, it is given in a slightly different form.

It can be shown that if the endpoints have a same absolute value, i.e. $|a| = |b|$, then the optimal weights at each point will be $\frac{1}{2}$. Moreover using (5.9), the optimal weights in the intervals $[0, 1]$, $[0, 2]$ and $[0, 10]$ can be computed straightforward, i.e. $w^* = \{\{0.59 \ 0.41\}, \{0.69 \ 0.31\} \text{ and } \{0.91 \ 0.09\}\}$, respectively.

Equation (5.7) describes how the $\text{Trace}\{\text{Var}(\hat{\beta})\}$ depends upon the parameter values a and b and on the weights w_1 and w_2 . It turns out that w_1^*

Table 5.1: Optimal weights for design intervals $[-1, 1]$, $[0, 1]$, $[0, 2]$ and $[0, 10]$ of first-order polynomial model ($p = 2$)

Scale	$[-1, 1]$		$[0, 1]$		$[0, 2]$		$[0, 10]$	
	d_{11}	ρ	w_1^*	w_2^*	w_1^*	w_2^*	w_1^*	w_2^*
0	0	0	0.50	0.50	0.59	0.41	0.69	0.31
		0.2	0.50	0.50	0.57	0.43	0.69	0.31
		0.6	0.53	0.47	0.57	0.43	0.69	0.31
		0.8	0.61	0.39	0.63	0.37	0.73	0.27
1	0	0	0.56	0.44	0.56	0.44	0.70	0.30
		0.2	0.57	0.43	0.57	0.43	0.71	0.29
		0.6	0.62	0.38	0.63	0.37	0.73	0.27
		0.8	0.69	0.31	0.70	0.30	0.78	0.22
10	0	0	0.76	0.24	0.71	0.29	0.83	0.17
		0.2	0.77	0.23	0.73	0.27	0.83	0.17
		0.6	0.80	0.20	0.79	0.21	0.86	0.14
		0.8	0.85	0.15	0.84	0.16	0.89	0.11

becomes large when e.g. b increases and a is kept fixed.

Algebraic solutions are difficult to obtain for models with $p > 2$ polynomial parameters, serially correlated errors, and random intercept terms. Therefore, in this paper, we will rely on results obtained by numerical methods.

Numerical approximation for the first- and second-order polynomial models with correlated errors

In Tables 5.1 and 5.2, the optimal weights for the serial correlations $\rho = 0, 0.2, 0.6, 0.8$ and random intercept variances $d_{11} = 0, 1, 10$ in the design intervals $[-1, 1]$, $[0, 1]$, $[0, 2]$ and $[0, 10]$ are summarized for the first- and second-order polynomial models, respectively.

In Table 5.1, the distribution of the optimal weights corresponding to the two A -optimal design points a and b for a first-order random intercept model ($p = 2$) with autocorrelated errors is shown. The values of the random intercept variance and serial correlation parameter are provided along with the scale.

Table 5.1 shows that the distribution of the optimal weights of first-order polynomial models ($p = 2$) is skewed to the right. In almost all intervals high values of ρ and/or high values of d_{11} are connected to a skewed distribution of weights. It is interesting to observe that in the interval $[0, 10]$, the skewed distribution of the optimal weights does not differ for different ρ s and slightly differs for different d_{11} s.

In Table 5.2, the distribution of the optimal weights for a second-order

Table 5.2: Optimal weights for design intervals $[-1, 1]$, $[0, 1]$, $[0, 2]$ and $[0, 10]$ of second-order polynomial model ($p = 3$)

Scale	d_{11}	ρ	$[-1, 1]$			$[0, 1]$			$[0, 2]$			$[0, 10]$		
			x_2^*	w_1^*	w_2^*	x_2^*	w_1^*	w_2^*	x_2^*	w_1^*	w_2^*	x_2^*	w_1^*	w_2^*
0	0	0	0.00	0.25	0.50	0.50	0.32	0.49	0.19	0.99	0.39	0.47	0.14	4.40
		0.2	0.00	0.26	0.48	0.46	0.39	0.39	0.22	0.92	0.43	0.43	0.14	4.40
		0.6	0.00	0.26	0.49	0.46	0.38	0.38	0.24	0.85	0.42	0.42	0.16	4.40
		0.8	0.00	0.23	0.54	0.46	0.38	0.38	0.24	0.81	0.43	0.43	0.14	4.20
1	0	0	0.00	0.27	0.48	0.48	0.40	0.40	0.20	0.99	0.42	0.42	0.16	4.60
		0.2	0.00	0.26	0.49	0.46	0.38	0.38	0.24	0.88	0.43	0.43	0.14	4.60
		0.6	0.00	0.23	0.54	0.46	0.38	0.38	0.24	0.81	0.43	0.43	0.14	4.40
		0.8	0.00	0.19	0.62	0.44	0.39	0.39	0.22	0.78	0.50	0.38	0.12	4.00
10	0	0	0.04	0.21	0.63	0.48	0.40	0.40	0.20	0.92	0.49	0.40	0.11	4.60
		0.2	0.00	0.18	0.66	0.44	0.40	0.40	0.20	0.88	0.52	0.37	0.11	4.60
		0.6	0.00	0.14	0.73	0.42	0.45	0.37	0.18	0.81	0.61	0.30	0.09	4.60
		0.8	0.00	0.11	0.78	0.44	0.55	0.29	0.16	0.81	0.69	0.23	0.08	4.40

polynomial model ($p = 3$) with random intercept and autocorrelated errors is shown along with the middle A -optimal design point x_2^* . Table 5.2 shows that the distribution of the optimal weights of second-order polynomial models ($p = 3$) is also skewed to the right. An exception occurs in the interval $[-1, 1]$, where high values of ρ and/or high values of d_{11} are connected to a peaked distribution of weights that is higher weights are assigned to the middle design point x_2^* . Notice that for the second-order polynomial model ($p = 3$) the skewed distribution of the optimal weights in the interval $[0, 10]$ differs slightly for different ρ s and moderately for different d_{11} s. Similar observations have been reported by Abt et.al. (1998) for intervals $[-a, a]$, where $a \geq 5$.

It also turns out that in the scale $[-1, 1]$, both first- and second-order polynomial models have two optimal designs, which means that the design $\xi_1^* = \{-1 \ 1; 0.44 \ 0.56\}$ is equivalent to the design $\xi_1^{*'} = \{-1 \ 1; 0.56 \ 0.44\}$ and the design $\xi_1^* = \{-1 \ 0 \ 1; 0.25 \ 0.48 \ 0.27\}$ is equivalent to $\xi_1^{*'} = \{-1 \ 0 \ 1; 0.27 \ 0.48 \ 0.25\}$. Thus, the optimal weights obtained in the interval $[-1, 1]$ can be swopped.

Last, the results of Tables 5.1 and 5.2 suggest a relation between the degree of skewness of the distribution of weights and the width of the interval $[a, b]$. See, for example, the A -optimal design for an uncorrelated first-order polynomial model having $d_{11} = 1$ in the interval $[0, 10]$: $\xi_1^* = \{0 \ 10; 0.92 \ 0.08\}$. This design allocates more weight on the first design point than the corresponding A -optimal design in the interval $[-1, 1]$, i.e. $\xi_1^* = \{-1 \ 1; 0.61 \ 0.39\}$. Similar results are observed for the second-order polynomial model.

Numerical approximation under equal weights restriction

In Figure 5.1, the location of the optimal time points x_{eq}^* (restriction of equal weights) and x_{un}^* (no restriction of equal weights) as a function of the autocorrelation parameter ρ in the range $[0, 0.8]$ and the random intercept variance $d_{11} = 1$ is displayed for the second-order polynomial model ($p = 3$). Each row of subplots (a) through (d) illustrates the behavior of the computed A -optimal designs in the intervals $[-1, 1]$, $[0, 1]$, $[0, 2]$ and $[0, 10]$, respectively.

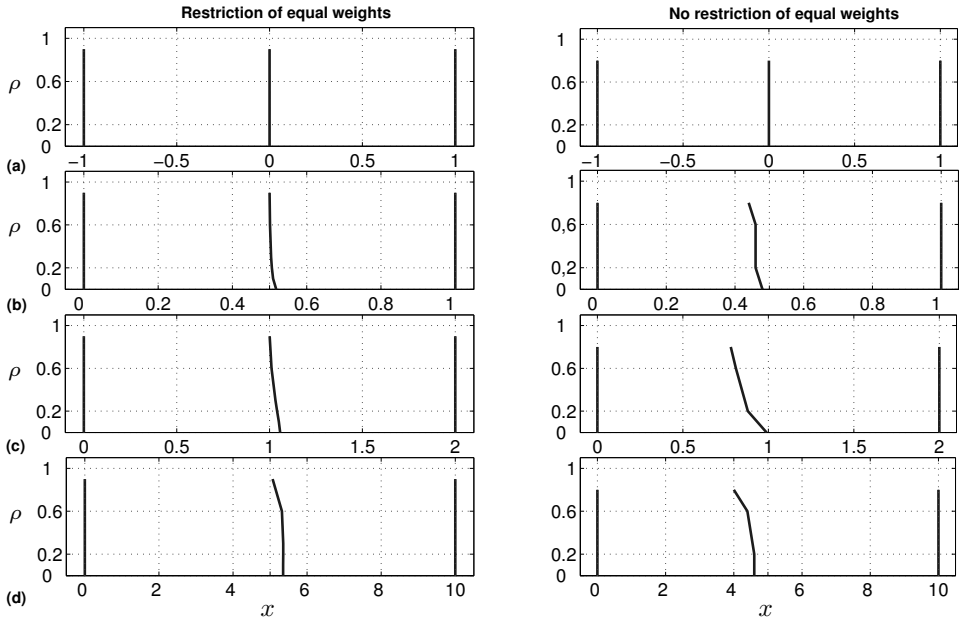
The left-column of subplots of this figure suggest that the location of the design points ξ_{eq}^* in one scale can be more or less obtained by a linear transformation from another scale. By comparing the left- with the right-subplots, it is seen that the location of the design points ξ_{eq}^* is slightly different from that of the design points ξ_{un}^* . Moreover, the design points ξ_{un}^* are clearly not obtainable by a linear transformation from another scale.

Plots similar to those presented in Figure 5.1 could have been plotted for random intercept variances $d_{11} = 0, 2, 3, 4$ and 10, as well. However, these plots are omitted because they are all similar to Figure 5.1. This means that the addition of a random intercept term does not seem to influence the location of the time points.

Moreover, similar plots (as those included in Figure 5.1) were generated for a first-order polynomial model ($p = 2$), i.e. varying the serial correlation and random intercept variances. However, we omit them as well because all these A-optimal designs are the same, i.e. a time point at each end of the design interval.

Notice that the A-optimal designs in the design interval $[0, 2]$ are equal to those reported by Moerbeek (2004), namely $\xi_{\text{eq}}^* = \{0, 2\}$ and $\xi_{\text{eq}}^* = \{0, 1.057, 2\}$, for the first- and second-degree polynomials, respectively.

Figure 5.1: Location of A-optimal design points for the second-order polynomial model ($p = 3$) with random intercept variance $d_{11} = 1$ in the intervals (a) $[-1, 1]$, (b) $[0, 1]$, (c) $[0, 2]$ and (d) $[0, 10]$.



5.4.2. Relative efficiency of designs applied to different intervals

As illustration of the numerical computations, to obtain relative efficiencies, we sketch here those for a first-order polynomial model ($p = 2$) with fixed-effects and uncorrelated errors. Using (5.7) in (5.6), we derive:

$$RE(\xi_1^*|\xi_2^*) = \frac{(c^2w_1^* + d^2w_2^* + 1)w_1w_2}{(c^2w_1 + d^2w_2 + 1)w_1^*w_2^*}. \quad (5.10)$$

So, if the design $\xi_1^* = \{0 \ 1; 0.59 \ 0.41\}$ which is optimal in the interval $[0, 1]$ is used in the interval $[0, 10]$, where $\xi_2^* = \{0 \ 10; 0.91 \ 0.09\}$ holds, the relative efficiency will be $RE = 0.703$. Similar values for other first-order polynomial models with uncorrelated errors can be easily computed using equation (5.10). It is, however, in general difficult to achieve similar algebraic expressions for models having random intercept terms with serially correlated errors and for higher order polynomials.

In the following, results for correlated errors with random intercept variances will be presented for the autocorrelation parameter $\rho = 0, 0.2, 0.6, 0.8$, the random intercept variance $d_{11} = 0, 1, 10$ and the intervals $[-1, 1]$, $[0, 1]$, $[0, 2]$ and $[0, 10]$.

Tables 5.3 and 5.4 represent the relative efficiencies RE of a chosen optimal design ξ_1^* (optimal in an interval $[a, b]$) when applied to another interval $[c, d]$ for first- and second-order polynomial models, respectively.

For example, the relative efficiency of using design $\xi_1^* = \{0 \ 2; 0.69 \ 0.31\}$ from Table 5.1, which is optimal for a fixed-effects model with $p = 2$ and uncorrelated errors in the interval $[0, 2]$, but which is suboptimal in the interval $[0, 10]$ where $\xi_2^* = \{0 \ 10; 0.91 \ 0.09\}$ (See Table 5.1) should be used instead, can be read from Table 5.3. This value is found by reading out the very first row (uncorrelated case: $\rho = 0$ and fixed-effects: $d_{11} = 0$) in the last column (ξ_2^* in $[0, 10]$) at the top-right corner (ξ_1^* in $[0, 2]$) of Table 5.3 and is equal to 0.816.

In both Tables 5.3 and 5.4, small to moderate (0.10 – 0.35) efficiency losses arise when designs optimal in a narrow scale ($[-1, 1]$, $[0, 1]$ or $[0, 2]$) are interchanged.

Moderate to large efficiency losses (0.25 – 0.65) are reported when designs optimal in a narrow scale are applied to a broad scale ($[0, 10]$). Designs that are optimal in $[0, 10]$ perform badly when applied to other intervals. In fact, relative efficiencies for the first- and second-order polynomial models can be as low as 0.328 and 0.302, respectively. A $RE = 0.328$ is found when the design $\xi_1^* = \{0 \ 10; 0.91 \ 0.09\}$ is used instead of $\xi_2^* = \{-1 \ 1; 0.50 \ 0.50\}$ for ρ and d_{11} equal to 0 and a $RE = 0.302$ is found when the design $\xi_1^* = \{0 \ 4.60 \ 10; 0.84 \ 0.13 \ 0.03\}$ is used instead of $\xi_2^* = \{-1 \ 0.04 \ 1; 0.21 \ 0.63 \ 0.16\}$ for $\rho = 0$ and $d_{11} = 10$.

Table 5.3 shows that the relative efficiency is related to the weights and

Table 5.3: Relative efficiencies $RE(\xi_1^*|\xi_2^*)$ obtained for the first-order polynomials ($p = 2$) models in the design intervals $[-1, 1]$, $[0, 1]$, $[0, 2]$ and $[0, 10]$

		ξ_2^*						ξ_2^*						
Scale	d_{11}	ρ	$[-1, 1]$	$[0, 1]$	$[0, 2]$	$[0, 10]$	Scale	d_{11}	ρ	$[-1, 1]$	$[0, 1]$	$[0, 2]$	$[0, 10]$	
ξ_1^*	$[-1, 1]$	0	0	1	0.971	0.873	0.599	$[0, 2]$	0	0	0.856	0.952	1	0.816
		0.2	1	0.983	0.874	0.599	0.2		0.856	0.934	1	0.816		
		0.6	1	0.992	0.904	0.634	0.6		0.896	0.941	1	0.816		
		0.8	1	0.998	0.947	0.726	0.8		0.929	0.951	1	0.859		
		1	0	1	1.000	0.922	0.655		1	0	0.919	0.919	1	0.812
		0.2	1	1.000	0.929	0.666	0.2		0.912	0.917	1	0.823		
	$[0, 1]$	0.6	1	1.000	0.948	0.723	0.6	0.945	0.951	1	0.844			
		0.8	1	1.000	0.965	0.797	0.8	0.958	0.962	1	0.893			
		10	0	1	0.985	0.975	0.820	10	0	0.970	0.903	1	0.893	
		0.2	1	0.990	0.980	0.831	0.2	0.974	0.931	1	0.893			
		0.6	1	0.999	0.981	0.862	0.6	0.974	0.960	1	0.923			
		0.8	1	0.999	0.989	0.910	0.8	0.981	0.976	1	0.950			
ξ_1^*	$[0, 1]$	0	0	0.968	1	0.960	0.703	$[0, 10]$	0	0	0.328	0.438	0.631	1
		0.2	0.981	1	0.945	0.680	0.2		0.328	0.410	0.627	1		
		0.6	0.994	1	0.942	0.680	0.6		0.365	0.420	0.634	1		
		0.8	0.998	1	0.962	0.748	0.8		0.472	0.510	0.707	1		
		1	0	1.000	1	0.922	0.655		1	0	0.367	0.367	0.612	1
		0.2	1.000	1	0.929	0.666	0.2		0.374	0.380	0.618	1		
	$[0, 1]$	0.6	1.000	1	0.956	0.734	0.6	0.454	0.466	0.681	1			
		0.8	1.000	1	0.972	0.808	0.8	0.589	0.598	0.785	1			
		10	0	0.986	1	0.937	0.767	10	0	0.501	0.375	0.687	1	
		0.2	0.993	1	0.951	0.788	0.2	0.511	0.417	0.696	1			
		0.6	0.999	1	0.974	0.851	0.6	0.610	0.568	0.781	1			
		0.8	1.000	1	0.984	0.900	0.8	0.746	0.729	0.879	1			

Table 5.4: Relative efficiencies $RE(\xi_1^*, \xi_2^*)$ obtained for the second-order polynomial model ($p = 3$) in the design intervals $[-1, 1]$, $[0, 1]$, $[0, 2]$ and $[0, 10]$

	ξ_2^*			ξ_2^*			ξ_2^*								
	Scale	d_{11}	ρ	$[-1, 1]$	$[0, 1]$	$[0, 2]$	$[0, 10]$	Scale	d_{11}	ρ	$[-1, 1]$	$[0, 1]$	$[0, 2]$	$[0, 10]$	
ξ_1^*	$[-1, 1]$	0	0	1	0.967	0.887	0.496	$[0, 2]$	0	0	0	0.878	0.969	1	0.733
			0.2	1	0.894	0.841	0.513			0.2	0.2	0.850	0.952	1	0.789
			0.6	1	0.837	0.788	0.510			0.6	0.6	0.883	0.946	1	0.772
			0.8	1	0.771	0.755	0.463			0.8	0.8	0.847	0.912	1	0.766
	1	1	0	1	0.864	0.811	0.531	1	1	0	0	0.906	0.986	1	0.729
			0.2	1	0.826	0.771	0.516			0.2	0.2	0.866	0.937	1	0.745
			0.6	1	0.762	0.742	0.479			0.6	0.6	0.849	0.918	1	0.735
			0.8	1	0.703	0.739	0.461			0.8	0.8	0.826	0.879	1	0.791
	10	10	0	1	0.644	0.751	0.622	10	10	0	0	0.832	0.886	1	0.665
			0.2	1	0.633	0.726	0.597			0.2	0.2	0.835	0.873	1	0.702
			0.6	1	0.657	0.756	0.568			0.6	0.6	0.838	0.872	1	0.803
			0.8	1	0.701	0.797	0.578			0.8	0.8	0.867	0.894	1	0.866
ξ_1^*	$[0, 1]$	0	0	0.967	1	0.972	0.619	$[0, 10]$	0	0	0	0.441	0.566	0.695	1
			0.2	0.933	1	0.967	0.721			0.2	0.2	0.428	0.506	0.677	1
			0.6	0.933	1	0.958	0.702			0.6	0.6	0.445	0.453	0.650	1
			0.8	0.881	1	0.950	0.681			0.8	0.8	0.517	0.460	0.702	1
	1	1	0	0.940	1	0.988	0.696	1	1	0	0	0.428	0.493	0.600	1
			0.2	0.938	1	0.951	0.662			0.2	0.2	0.435	0.450	0.614	1
			0.6	0.882	1	0.946	0.653			0.6	0.6	0.506	0.448	0.673	1
			0.8	0.805	1	0.931	0.634			0.8	0.8	0.545	0.437	0.720	1
	10	10	0	0.807	1	0.943	0.547	10	10	0	0	0.462	0.302	0.521	1
			0.2	0.778	1	0.926	0.547			0.2	0.2	0.494	0.324	0.563	1
			0.6	0.745	1	0.896	0.605			0.6	0.6	0.604	0.408	0.689	1
			0.8	0.785	1	0.913	0.703			0.8	0.8	0.717	0.495	0.786	1

that this relation seems to be non-linear.

In the case of narrow scales, i.e. $[-1, 1]$, $[0, 1]$ and $[0, 2]$, the serial correlation parameter ρ seems to account for most of the efficiency loss. Higher serial correlations imply usually lower RE 's. It also looks like the presence of a random intercept variance influences this drop more than the value of d_{11} itself.

Notice that the behavior of the relative efficiency for the second-order ($p = 3$) polynomial models resembles that of the relative efficiencies of $p = 2$ polynomial models. The RE s for $p = 3$ are displayed in Table 5.4.

5.4.3. Relative efficiency of designs applied under equal weight restriction

As illustration of the numerical computations of the RE 's in this context, the case of a first-order polynomial model ($p = 2$) with fixed-effects is sketched here. Using (5.7) in (5.6), we derive:

$$RE(\xi_{eq}^*|\xi_{un}^*) = \frac{1}{2w_1^*} \frac{(b^2 + 1)}{(a^2 + b^2 + 2)} + \frac{1}{2w_2^*} \frac{(a^2 + 1)}{(a^2 + b^2 + 2)}, \quad (5.11)$$

where w_1^* and w_2^* are the optimal unequal weights of the design ξ_{un}^* . These optimal weights are obtained by using the method of all possible combinations. The interval parameters a and b are fixed at the endpoints.

So, if the optimal design $\xi_{eq}^* = \{0 \ 1; 0.50 \ 0.50\}$ is used in the interval $[0, 1]$ under the constraint of equal weights instead of the A-optimal design $\xi_{un}^* = \{0 \ 1; 0.59 \ 0.41\}$, the relative efficiency will be $RE = 0.971$.

Notice that the relative efficiency of choosing an A-optimal design with equal weights instead of its unequal weights counterpart will drop rather quickly whenever a is fixed and b is allowed to increase. The reason is that if b is large and w_1^* approaches to 1 then the first term of equation (5.11) becomes very large and the second term which is connected to w_2^* adds relatively less to the relative efficiency.

The relative efficiencies when an optimal design ξ_{eq}^* having equal weights is used instead of its unequal weights counterpart ξ_{un}^* are shown in Table 5.5 for the first- and second-order polynomial models. This table follows the convention of Table 5.4, i.e., the first-, second- and third-level correspond to the scale or design interval, the random intercept variance d_{11} and the serial correlation ρ , respectively.

This table shows that the $RE(\xi_{eq}^*|\xi_{un}^*)$ also depends on the order of the polynomial, the chosen interval and the variance-covariance structure. It is observed that usually when the serial correlation ρ and/or the random in-

Table 5.5: Relative efficiencies $RE(\xi_{eq}^*|\xi_{un}^*)$ for the first- and second-order polynomial ($p = 2, 3$) model given that a ξ_{eq}^* is chosen instead of the ξ_{un}^* for the design intervals $[-1, 1]$, $[0, 1]$, $[0, 2]$ and $[0, 10]$.

d_{11}	ρ	$p = 2$				$p = 3$			
		ξ_{un}^* $[-1, 1]$	$[0, 1]$	$[0, 2]$	$[0, 10]$	ξ_{un}^* $[-1, 1]$	$[0, 1]$	$[0, 2]$	$[0, 10]$
0	0	1	0.971	0.873	0.599	0.898	0.895	0.869	0.625
	0.2	1	0.983	0.874	0.599	0.919	0.954	0.878	0.625
	0.6	0.996	0.979	0.871	0.599	0.912	0.971	0.885	0.620
	0.8	0.9456	0.937	0.831	0.598	0.841	0.970	0.875	0.599
ξ_{eq}^*	1	0	0.984	0.984	0.857	0.586	0.921	0.941	0.883
	0.2	0	0.981	0.979	0.854	0.586	0.911	0.960	0.881
	0.6	0	0.943	0.937	0.820	0.586	0.841	0.967	0.872
	0.8	0	0.870	0.865	0.764	0.583	0.741	0.958	0.837
10	0	0	0.782	0.854	0.699	0.542	0.724	0.930	0.829
	0.2	0	0.777	0.828	0.696	0.542	0.693	0.941	0.809
	0.6	0	0.731	0.749	0.663	0.541	0.601	0.911	0.723
	0.8	0	0.677	0.683	0.625	0.539	0.528	0.818	0.628

tercept variance d_{11} increase, the relative efficiency decreases. Furthermore, the loss of efficiency is more pronounced when the endpoint b becomes large.

The behavior of the relative efficiency for the first-order polynomial looks is similar to that of the second-order polynomial model but is displaced slightly downwards. The RE s of the first- and second-order polynomial models can reach values up to 0.534 and 0.439, respectively.

The $RE(\xi_{eq}^*|\xi_{un}^*)$ also depends on the design interval. For example, optimal designs for a second-order polynomial model (with $d_{11} = 10$) in the interval $[0, 10]$ have lower RE 's than those in the interval $[-1, 1]$.

Designs which are optimal for uncorrelated models are more efficient than those for highly correlated models. See, for example, the case of a design for a first-order polynomial model with a random intercept $d_{11} = 1$ in the scale $[-1, 1]$. In Table 5.5, when $\rho = 0$, the relative efficiency is 0.984 while when $\rho = 0.8$ the RE value drops to 0.870. Notice also that as b becomes larger than a , the difference between the RE of uncorrelated and highly correlated designs fades out.

Finally, the RE depends on the value of the random intercept variance d_{11} . Especially, when the underlying model is a second-order polynomial, the designs for fixed-effects models are more efficient than those for random intercept models with large d_{11} .

5.5. Discussion and conclusion

This paper addresses the question: *Are A-optimal designs for random intercept models with serially correlated errors robust against changes on their design scales?*

Our results show that designs optimal in scales with narrow intervals have moderate losses of efficiency when scales are interchanged. Large losses of efficiency are obtained when designs optimal in broad interval scales are applied to narrow interval scales and viceversa. Moreover, in both cases, the relative efficiency depends on the order of the underlying polynomial model and on the model variance-covariance structure. Also, worth to mention is that the relative efficiency RE is related to its corresponding optimal and suboptimal weights in a non-linear way and that the same non-linear behavior is observed for the serial correlation parameter and the random intercept variance.

Furthermore, similar non-linear behavior holds for the relative efficiency of using a design with equal weights instead of its unequal weights counterpart. Here, the RE s are the lowest for designs optimal in broad intervals and optimal designs for uncorrelated models are more efficient than those for highly correlated models. Notice also that for the second-order polynomial, the obtained RE s also account for the changes of the optimal middle point.

Our main conclusion is that A -optimal designs depend strongly on their design scale and on their optimal weights. Suboptimal designs arise when the scale of the design interval is broad and when the restriction of equal weights is introduced. Both situations can lead to non-negligible losses of efficiency.

Since longitudinal studies rely on time scales and on the allocation of subjects to specific design points, the use of A -optimal designs in these studies is definitely not advisable. Instead the practitioner is advised to use other optimality criteria. In case that the A -optimality criterion remains being preferred by the practitioner, he/she should compute the optimal designs for his/her working design scale. Notice though that if the range of the interval is large some parameters will be inaccurately estimated and the reliability of the study will decrease. For example, if a first-order polynomial model is used in a longitudinal study, then the practitioner may find that the optimal designs for the scale $[0, 1]$ (1-year) and the optimal designs for the scale $[0, 12]$ (12-months) can lead to different conclusions. The explanation is that the weights of the optimal design having the scale $[0, 12]$ will be concentrated at the point 0, which means that the slope of this model will be badly estimated.

It is known that D -optimal designs are scale invariant and although they do not always have equal weights, the optimal location of the design points for different order of polynomial models and for different variance-covariance structures have been published in Tan and Berger (1999); Ouwens et.al. (2002); Pukelsheim (1993); Wong (1994) and Moerbeek (2004) among others. A linear transformation can be used to change a design scale into another design scale, and thus provide the optimal location of the design points on the design scale of interest.

Last, although we report the first- and second-order random intercept polynomial models only, it is expected that the profile of the relative efficiency of optimal designs connected to higher-order polynomial models will show a similar non-linear profile. However, this profile will be slightly shifted downwards. Similar behavior of the RE s also holds for polynomial models with random slope variance and serially correlated errors.

CHAPTER 6

ODMIXED: A tool to obtain optimal designs for
heterogeneous longitudinal studies with dropout ¹

¹*Submitted for publication*

Abstract

ODMIXED is a computer program to obtain optimal designs for linear mixed models of longitudinal studies. These designs account for heterogeneous correlated errors and for data with dropout. Designs are compared by using relative efficiencies, e.g. between a D-optimal design for homogeneous data and another for heterogeneous data or between a D-optimal design for complete data against another that optimizes designs when data is missing at random. Two examples are worked out to illustrate how researchers may profit of optimal design theory at the planning stage of longitudinal studies.

Keywords: D-optimal designs; Dropout; Heterogeneous autocorrelation; Linear mixed models; Longitudinal data; Relative efficiency.

6.1. Introduction

At the planning stage of longitudinal studies, the allocation of the resources (time, subjects and/or money) is a critical issue. To collect longitudinal data, it is common practice to use equally-spaced designs since these designs are model-free. However, these designs can be suboptimal when compared with their optimal counterparts (Tekle, Tan and Berger, 2008a) or when data is missing (Ortega, Tan and Berger, 2008). Both studies showed that at the planning stage of longitudinal data, researchers may profit, substantially, from optimal design theory.

Optimal design of experiments give the lowest estimators variance such that the estimators have high precision. However, these designs are model-dependent, i.e., the researcher must have prior knowledge about the underlying model fitting the longitudinal data. These data are usually correlated, can have heterogeneous variances and/or can be affected by dropout. The most suitable model to fit data with correlated and unbalanced data structures is the linear mixed models (Laird and Ware, 1982).

In this paper, we present a computer program ODMIXED that computes optimal designs for complete data, for data suffering from dropout and for data having heterogeneous error structures. ODMIXED is made in MATLAB. MATLAB is chosen because of its flexible plotting capabilities, robust optimization algorithms and steadily growing number of toolboxes, not to mention the fact that it is been steadily introduced in biomedical and health sciences applications.

To our knowledge this is the first program that computes optimal designs for heterogeneous longitudinal data and data missing at random. Notice that the optimal designs with complete data and homogenous error structure can be matched with those obtained for one-cohort using POLS. This is an interactive program implemented in MATLAB that allows to compute D-optimal designs for different polynomial models with mixed effects for a set of different number of cohorts (Tekle, Tan and Berger, 2009).

This paper is organized as follows. In section 6.2, two motivating examples will be introduced. In section 6.3, the design problem for the data with dropout and heterogeneous error variance structure will be briefly discussed and the relative efficiency will be introduced as a measure to compare two or more designs. In section 6.4, the graphical user interphase will be presented. Finally, some conclusions and recommendations are provided in section 6.5.

Table 6.1: Dropout pattern of the psycho-and pharmacotherapy follow-up study for cocaine addicts (Carroll et.al., 1994).

	months				
	0	1	3	6	12
Four Interviews		25	25	25	25
Three Interviews		25	25	25	
Two Interviews		16	16		
One Interview		31			
Subjects (n_j)	121	97	66	50	25
index(j)	1	2	3	4	5
Response probability $p(x_j)$	1	0.80	0.54	0.41	0.21

6.2. Motivating examples

To illustrate the use of the ODMIXED program, we consider the cases of a randomized clinical trial and a longitudinal study.

6.2.1. Example 1: Randomized clinical trial on psycho- and pharmacotherapy

In clinical trials, complete data sets are rare. Patients can dropout during the follow-up and tracking them may become a very costly task. We consider the data from a 1-year follow-up of a randomized clinical trial of psychotherapy (cognitive-behavioral relapse prevention or clinical management) and pharmacotherapy (desipramine or placebo) for ambulatory cocaine addicts (Carroll et.al., 1994; Nich et.al., 1997).

Patients were randomly assigned to one of four treatments. Outcomes were measured using urine toxicology screens and the Addiction Severity Index. A total of 121 patients was scheduled to be followed 1, 3, 6 and 12 months after termination of the treatment. The first month after treatment, the response percentage was 80%. Only 54% of the patients were measured at 1 and 3 months; 41% were measured at 1, 3 and 6 months; and as much as 21% of the subjects finished the follow-up. This dropout process is described in Table 6.1 in terms of a response probability function and the number of patients that are available at each design point.

At the design stage of clinical trials, it is important to foresee the occurrence of dropout, i.e. data missing at random. However, although researchers may have reasonable guesses about this function, they usually have no knowledge of its exact form.

We use this example to know how much efficiency is lost during the collection of longitudinal data when an optimal design, that incorrectly assumes complete data, is used instead of that for data with dropout .

6.2.2. Example 2: Longitudinal study of elderly chronic care populations

In gerontological studies, it is important to collect data as efficiently as possible in order to describe the decline of mental state of elderly chronic care populations appropriately (Petkova and Teresi, 2002; Holmes et.al., 2000). In this study the decline of functional limitations is compared over time of 186 demented and non-demented elderly residents in long-term facilities. Linear mixed models were used to describe the data. Several polynomial functions were used to describe trends over time. A serial autocorrelation structure was used to describe the within subject time-structured data.

The data reported was collected using equally-spaced design points. Three time points were considered: baseline, first follow-up and second follow-up. The planned interval among time points was 6 months, i.e., observations correspond to 0, 6 and 12 months, respectively.

Both studies (Petkova and Teresi, 2002; Holmes et.al., 2000) considered the case when variances of the outcome were not the same at different observation times, i.e. there is a heterogeneous autoregressive error covariance structure. In first instance, it is usually assumed that the data can be fitted by a linear mixed effects model with homogeneous error structure. In many cases, however, the variances of the measurements change over time and the data might be better fitted by a heterogeneous error covariance structure. We use this example to study the consequence of incorrectly choosing an optimal design for a homogeneous error covariance structure.

6.3. Longitudinal data and linear mixed models

In longitudinal studies, the conditions for continuous responses \mathbf{y}_i of n subjects must be planned well ahead the collection of the data. A typical model for longitudinal data is the linear mixed models. It is defined as

$$\mathbf{y}_i = X_i\boldsymbol{\beta} + Z_i\mathbf{b}_i + \mathbf{e}_i, \quad (6.1)$$

where $i = 1, \dots, n$ are the subjects participating in the biomedical or social longitudinal study. The observations \mathbf{y}_i are taken, under conditions x_j , from the j^{th} time-point, i.e. a date in the planning scheme of the longitudinal study.

The model (6.1) contains three components: a fixed-effects term $X_i\beta$, a random-effects $Z_i\mathbf{b}_i$ and a random error \mathbf{e}_i . The matrix X_i is a known $q \times p$ design matrix. The vector β denotes a $p \times 1$ vector of unknown fixed parameters; p and q represent the number of the fixed parameters of the polynomial regression and the number of time points, respectively. Let \mathbf{b}_i denote a $k \times 1$ vector of unknown random-effects parameters and Z_i be a known $q \times k$ design matrix. \mathbf{b}_i represents how the i^{th} subject deviates from the average population and it is assumed to be normally distributed with mean $\mathbf{0}$ and a $k \times k$ covariance matrix D . Notice that we will restrict ourselves to $k = 0, 1, 2$ meaning that three cases are considered: fixed-effects, random-intercept and random intercept- and -slope models.

The random error of each subject is represented by \mathbf{e}_i . This error is assumed to be normally distributed with mean $\mathbf{0}$ and covariance $\Sigma_i^{1/2}\Psi_i\Sigma_i^{1/2}$. Σ_i is a diagonal matrix with variances $(\sigma_1^2, \dots, \sigma_{q_i}^2)$ on the main diagonal. We allow that observations at different time-points are (serially) correlated. This dependence is modeled through the within-subject correlation matrix Ψ_i , i.e. a positive-definite autocorrelation matrix depending on the parameter ρ . The parameter ρ denotes the autocorrelation between the j^{th} and j'^{th} time points x_j and $x_{j'}$, respectively. The elements of the autocorrelation matrix are given by $\psi_{jj'} = \rho^{\text{abs}(x_j - x_{j'})}$, where $j, j' = 1, 2, \dots, q$. Notice that the time points x_j are not necessarily equally spaced. Note that \mathbf{b}_i and \mathbf{e}_i are independent from each other.

Two cases are distinguished. First, errors \mathbf{e}_i are homogeneous among different time points x_j and $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_q^2 = \sigma^2$. So, $\Sigma_i^{1/2}\Psi_i\Sigma_i^{1/2}$ simplifies to $\sigma^2\Psi_i$. Second, the error variance function is heterogeneous, i.e. $\Sigma_i^{1/2}\Psi_i\Sigma_i^{1/2}$; where the diagonal elements of Σ_i are all unequal in value.

Notice that from now on the subindex i will be omitted since the measurements of each subject i are performed at the same time-points.

6.3.1. Optimal designs and response probability function

The exact design ξ for the linear mixed model (6.1) is defined as:

$$\xi = \left\{ \begin{array}{cccccc} x_1 & x_2 & \dots & x_j & \dots & x_q \\ n_1 & n_2 & \dots & n_j & \dots & n_q \end{array} \right\}, \quad \text{where } j = 1, \dots, q, \quad (6.2)$$

At the first design point x_1 of the study, the probability of response is $p(x_1) = 1$ and the number of subjects responding at x_1 is equal to n_1 . At the j^{th} design point with response probability $p(x_j)$, the expected number of subjects $n_j = n_1 p(x_j)$. Finally, at the last design point x_q , the number of subjects n_q represents the number of subjects completing the experiment.

For $j = q$, n_q is the number of subjects responding at all q design points. We will assume that at least one subject is observed at this last point, i.e., $n_q \geq 1$. It is also assumed that $p(x_j)$ is a monotonically decreasing function with $p(x_1) \geq p(x_2) \geq \dots \geq p(x_q)$ where $p(x_1) > p(x_q)$.

Notice that a response probability $p(x_j)$ at a design point x_j is complementary to a dropout process $1 - p(x_j)$ at the same design point x_j and that an important assumption considered all along this paper is that the dropout occurs through a noninformative mechanism, i.e., responses are missing at random (MAR), see e.g. Little and Rubin (1987).

For complete data, the number of subjects responding to each x_j is assumed equal, i.e., $p(x_j) = 1$ for all j . But, if a dropout process arises, the number of subjects responding at x_j will (monotonically) decrease. So, the probability of obtaining data at design point x_j will depend upon each design point. The values of the design time points x_j and the response probability at those points are confined in a given design space \mathcal{X} and probability range $0 < p(x_j) \leq 1$.

Our computer program ODMIXED can handle two polynomial response functions, namely, a linear function $p_{\text{lin}}(x_j) = a_0 + a_1 x_j$ and a quadratic function $p_{\text{quad}}(x_j) = a_0 + a_1 x_j + a_2 x_j^2$. Further details are given in Ortega, Tan and Berger (2008).

6.3.2. Asymptotic covariance matrix

We introduce the super-index $[j]$ to group subjects having j measurements with design matrices $X^{[j]}$ for the fixed-effects and $Z^{[j]}$ for the random-effects and connected to the response probability $p(x_j)$.

In general, if the number of subjects having all responses to all q design points is denoted by $m_q = n_q$ and the number of subjects with only $j < q$ responses is denoted by the difference $m_j = n_j - n_{j+1}$, then the number of subjects m_j having j responses is given as:

$$m_j = \begin{cases} n_q & \text{if } j = q, \\ n_j - n_{j+1} & \text{if } j < q. \end{cases} \quad (6.3)$$

The design matrices $X^{[j]}$ and $Z^{[j]}$ ($j = 1, \dots, q$) have size $j \times p$ and $j \times k$, respectively. The linear mixed model (6.1) for data with dropout and heterogeneous error implies a marginal model with mean $X_i \beta$ and covariance matrix:

$$V^{[j]} = Z^{[j]} D Z^{[j]'} + \Sigma^{[j]1/2} \Psi^{[j]} \Sigma^{[j]1/2}. \quad (6.4)$$

Notice that the last term of (6.4) reduces to $\sigma^2 \Psi^{[j]}$ when the error is homogeneous.

The matrix $V^{[j]}$ is connected to the asymptotic covariance matrix of the best linear unbiased estimator $\hat{\beta}$ and since m_j subjects share identical $X^{[j]}$ and $Z^{[j]}$, the asymptotic covariance matrix of the best linear unbiased estimator $\hat{\beta}$ can be estimated by:

$$\text{Var}(\hat{\beta}) = \left(\sum_{j=1}^q m_j X'^{[j]} V^{[j]-1} X^{[j]} \right)^{-1}, \quad (6.5)$$

where m_j follows (6.3) when data has dropout and $m_j = n_q$ for all j , when data is complete.

The asymptotic covariance matrix (6.5) is equal to the inverse of the Fisher information matrix. Consequently, small valued asymptotic covariance matrices of the parameter estimators correspond to estimators with high precision.

6.3.3. Generalized variance of the estimated parameters

In this paper, we will focus on the determinant (or D -optimality) criterion. This criterion is scale invariant and has a natural interpretation in terms of the volume of the simultaneous confidence ellipsoid of the fixed parameters (Atkinson and Donev, 1996).

A design ξ^* is called the D -optimal design of a model if it satisfies the inequality:

$$\text{Det}[\text{Var}(\hat{\beta}|\xi^*)] \leq \text{Det}[\text{Var}(\hat{\beta}|\xi)] \quad \forall \xi \in \mathcal{X}, \quad (6.6)$$

where \mathcal{X} is the design space and the determinant of the asymptotic variance covariance matrix is known as the generalized variance of the estimated parameters $\hat{\beta}$.

In this paper, we focuss on two polynomial error variance-covariance structures, namely, a linear function $\Sigma^{[j]}(x_j) = a_0 + a_1 x_j$ and a quadratic function $\Sigma^{[j]}(x_j) = a_0 + a_1 x_j + a_2 x_j^2$, where $a_0 > 0$ and $a_1, a_2 \in \mathcal{R}$. The latter function can have either a convex or a concave form.

Notice that the combined case of data with dropout and heterogeneous error variance structure is not considered in this paper.

6.3.4. Relative efficiency to compare designs

A measure to compare designs is the relative efficiency RE . In general, it is defined as

$$RE(\xi|\xi^*) = \left[\frac{\text{Det}\{\text{Var}(\hat{\beta}|\xi^*)\}}{\text{Det}\{\text{Var}(\hat{\beta}|\xi)\}} \right]^{\frac{1}{p}}. \quad (6.7)$$

Since, we study the effect of an incorrect specification of the dropout process (or the heterogeneity of the error variance) on the relative efficiency of D -optimal designs, it is assumed that the underlying model is fixed. So, both optimal designs have the same variance-covariance structure, the same number of time points q and the same number of parameters p .

The RE given by (6.7) is a ratio that measures the efficiency change when a study is planned with a suboptimal design (for complete data or for homogeneous error variance), while instead the optimal design (for data with dropout or for heterogeneous error variance) should have been used. Hence, the numerator of (6.7) has a smaller variance than the denominator and it is always smaller than 1.

Furthermore, notice that the ratio RE is proportional to the design size, irrespective of the dimension of the model, since it is raised to the power $1/p$, where p is the number of the fixed parameters of the models (Atkinson and Donev, 1996; Pukelsheim, 1993). A relative efficiency $RE = \frac{1}{2}$ means that the optimal design for a model with incorrectly assumed structure (complete data or homogeneous error variance) is suboptimal with respect to the optimal design for its true structure (data with dropout or heterogeneous error variance). Consequently, the sample size of (used) suboptimal design is twice larger than the size of the true optimal design. For more details about linear mixed models when data has dropout, the reader is referred to Ortega, Tan and Berger (2008).

6.3.5. Algorithm

D -optimal designs are found numerically by optimizing the generalized variance of the parameter estimators (6.6) in an arbitrary interval $[lb, ub]$, where lb and ub stand for the lower and upper bounds.

The program uses the implemented `FMINSEARCH` and `FMINCON` functions of MATLAB version 7.0.1(R14). These functions perform (un)constrained nonlinear optimization. The first function executes a direct search method while the second uses a sequential quadratic algorithm. The basic difference between these two functions is that the first does not use numerical

or analytic gradients, while the second is a gradient-based method. For more details, see (Matlab, 2004, Optimization Toolbox) and Molenberghs and Kenward (2007).

We compute two situations of relative efficiencies using (6.7). The first situation compares a design which is optimal under complete data assumption with that when dropout occurs. The second situation discriminates between a design which is optimal for a homogeneous error structure with a design for a heterogeneous error structure.

ODMIXED has run in a desktop computer with dubble INTEL processor x86, Family 6, Model 15. This computer has a speed of 1595 MHz. and required about a second to compute an optimal design. A plot alike Figure 6.7(Left) or (Right), needed about ten seconds to be obtained. The computational time for the relative efficiency plots was about 60 seconds for the example 2 on dementia patients, i.e. 36 variance-covariance structures. Notice that the generation of the plots and the saving of the results must also be accounted.

6.4. ODMixed

We devote this section to introduce the graphical user interphase of the computer program ODMIXED. This program has research and teaching purposes. It deals with two common situations, optimal designs for longitudinal data with dropout and optimal designs for (heterogeneous) autocorrelated errors.

The program is started by entering the command ODMIXED at the MATLAB prompt. (Notice that the current directory in the MATLAB desktop must be that containing the ODMIXED program.) Next, the layout of the program is displayed as in Figure 6.1. The program has four panels. The first panel allows the user to 'Introduce the parameters of the model' and the second panel is a selection panel which permits to 'Input variance-covariance data'. These panels are highlighted by a grey-colored background indicating that the user can change the 'Default Values'.

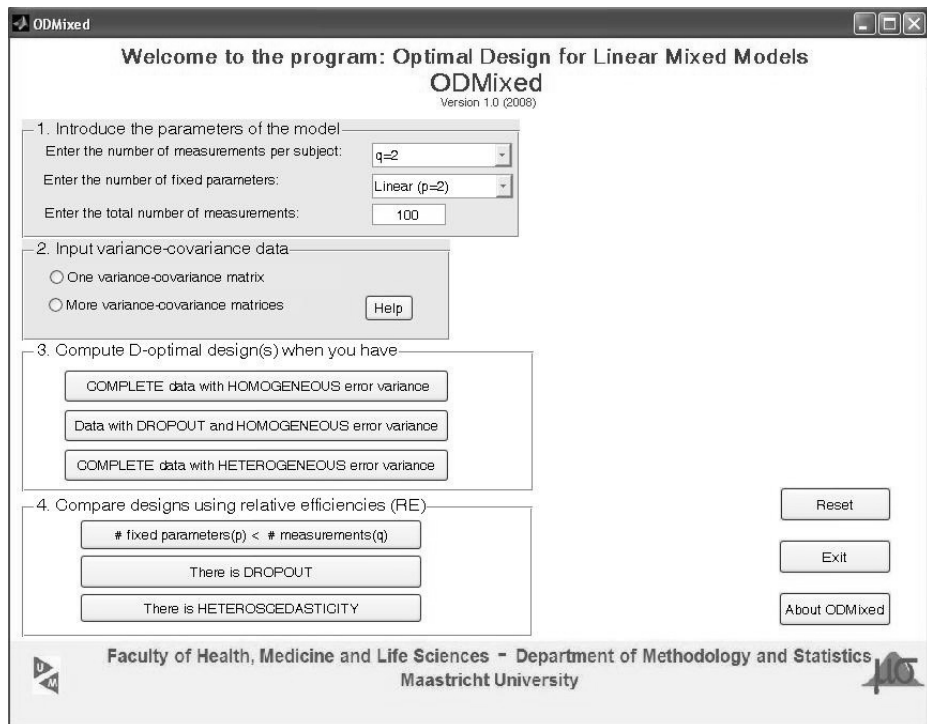
In the first panel, the user has two popup menus, one for the number of measurements per subject, $q = 2, 3, 4, 5$, and another for the number of parameters, i.e. to chose whether the model is Linear ($p = 2$); Quadratic ($p = 3$); Cubic ($p = 4$). To enter the total number of measurements (N) the user has to use the edit box (white-colored box).

In the second panel, the user has two options to introduce the variance-covariance structure: 'One variance-covariance matrix' or 'More variance-covariance matrices'.

The third panel of Figure 6.1 is meant to 'Compute optimal designs when you have' three situations. Each situation is connected to one of the three push buttons on this panel. Press the first push button to compute optimal designs when there is 'COMPLETE data with HOMOGENEOUS error variance' structure; press the second button to obtain optimal designs for 'Data with DROPOUT and HOMOGENEOUS error variance', or push on the third button to find designs for 'COMPLETE data with HETEROGENEOUS error variance' structure.

In the fourth panel, the user has the possibility to 'Compare alternative designs by using relative efficiencies (RE)'. The first push button allow to compute the relative efficiency of a D -optimal design when the ' $\#$ fixed parameters (p) < $\#$ repeated measurement (q)'. It means that if the user has $q = 4$ time points and his/her guess is that the underlying model is linear ($p = 2$), the program will provide the RE of designs optimal for quadratic and cubic linear mixed models. The second and third push buttons permit to compute the RE when 'There is DROPOUT' and 'There is

Figure 6.1: Layout of the main menu of ODMIXED



HETEROSCEDASTICITY', respectively.

Located at the right of the screen the user finds three standing alone push buttons. The button 'Reset' is to clear up all the global and local variables generated by the program, close down any screen that might be open and re-start the program. The button 'Exit' exits the program ODMIXED after having asked the user for a confirmation to leave the program. Next, it closes down all connected screens and clears up all global and local variables. The last button 'About ODMIXED' provides background information on the program.

Notice that when the user selects the option 'One variance-covariance matrix' in the input menu, the program displays Figure 6.2. In this screen the 'Default Values' of the variance-covariance parameters and the design space are provided. All (edit)boxes can be modified by the user. Only numerical values are allowed.

Figure 6.2: Layout of the input values menu of ODMIXED

INPUT VALUES

1.1 Introduce the parameters of the variance covariance matrix

Enter the values of the serial correlation range: 0.001 --- 0.9

Enter the value of the random intercept variance: 0

Enter the value of the random slope variance: 0

Enter the correlation between the random parameters: 0

Enter the values of the time interval range: -1 --- 1

1.2 Choose how to enter more data

Add another variance-covariance matrix

Open file with variance-covariances

Help

Save

Reset

Cancel

Faculty of Health, Medicine and Life Sciences - Department of Methodology and Statistics
Maastricht University

In case that the user is interested on studying more than one variance-covariance structure, he/she will have to use the panel 'Choose how to enter

more data'. In that panel, there are two push buttons. The first permits to 'Add another variance-covariance matrix'. A maximum of two variance-covariance structures is allowed via this option. Should the user have more variance-covariances, the user can better select a file containing the parameters of interest using the option 'Open file of variance-covariance matrices'. The 'Help' button located next to the button 'Open file...' explains the user the order in which the data should be uploaded, i.e., the data columns should be ordered as follow. First, the random intercept variance; second, the random slope variance; third, the correlation between random parameters; fourth and fifth, the lower- and upper-bounds of the serial correlation parameter; sixth and seventh, the lower and upper bounds of the time range. The user confirms the data by clicking on the button 'Open'. Notice that the format of the file is '(*.txt)'.

Once the user has introduced the input values, he/she must push the button 'Save'. In case that the user made a mistake, he can use the buttons 'Reset', i.e. put the values of the variance-covariance matrix and models parameters to their 'Default' status. Last, it can also occur that the user may leave this screen and return to the main menu. In that case, press the button 'Cancel'.

6.4.1. Example 1 (Revisited): Randomized clinical trial on psycho- and pharmacotherapy

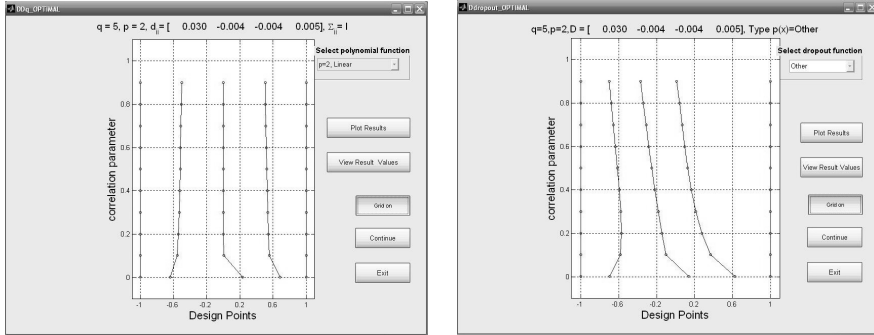
Consider the example provided in section 6.2. The D -optimal designs and relative efficiencies (RE 's) are obtained as follow.

First, fill out the input for the model given in Figure 6.1, i.e. select $q = 5$, since the longitudinal data has a five time-points. Then, select $p = 2$ since data are fitted by a first-degree polynomial. Next, fill in the value for N assuming complete data, i.e. 121 patients/time-point \times 5 time-points = 605 planned measurements. The measurements on Addiction Severity Index (ASI) of 121 cocaine addicts were reported.

Thereafter, in the input values screen (Figure 6.2,) adjust the values of the variance-covariance structure of the input screen given in Figure 6.2 to those of the example. The estimated variance-covariance matrix D of the random parameters was $D = (0.030 \quad -0.004; -0.004 \quad 0.005)$. The estimated correlation between random intercept and slope variances was $\rho^* = -0.328$. Notice that we can either use the time scale $[-1, 1]$ or the scale $[0, 12]$.

Notice that there are two ways to obtain the D -optimal designs for complete data. Pushing the button 'COMPLETE data with HOMOGENEOUS error structure' or pressing the button 'Data with DROPOUT and HOMO-

Figure 6.3: Left: D -optimal design assuming that data is complete. Right: D -optimal design when data has the probability function $p(x_j^d) = 0.36 - 0.37x_j + 0.23x_j^2$



GENEOUS error variance'. We chose the first option. A screen containing a menu panel and five push-buttons is displayed. In the menu panel the user must 'Select the degree of the polynomial model'. Then, by pressing the button 'Plot Results', the results will be displayed, as in Figure 6.3(Left).

The other option to obtain the same results is using the button for data with dropout and then select a response probability function that is 'Constant' and press 'Plot Results'.

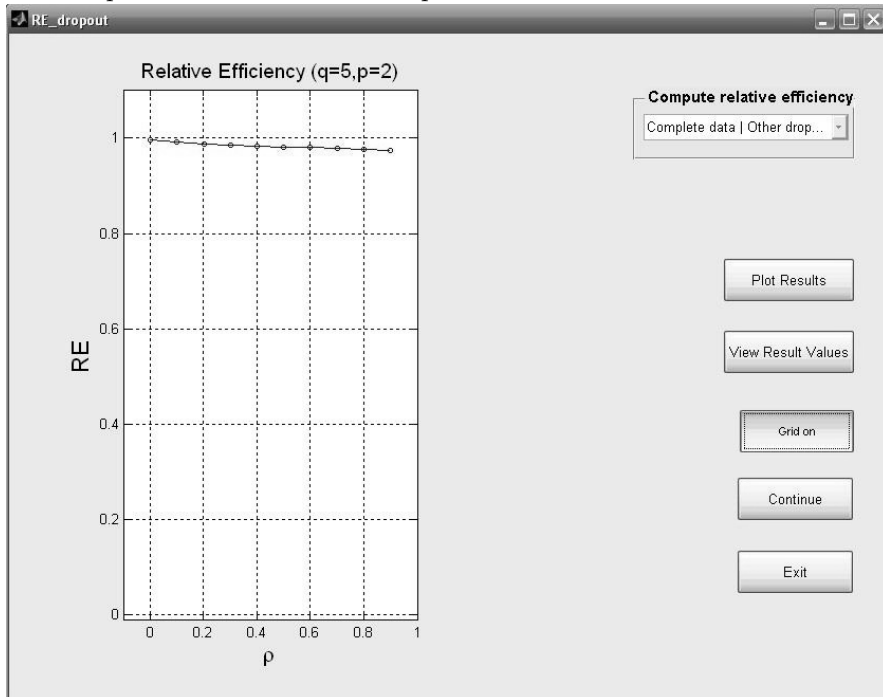
Notice that from both screens, the user can compute and view the results of the optimization. Graphical results will be given by pushing the button 'Plot results' and the text information by using the push button 'View Result Values'. The plots of the generated designs show the location of the time points against the value of the serial correlation. A title is displayed at the top of each plot. This title identifies the number of parameters, the number of design points and the random variance-covariance structure used during the optimization. If a list of variance-covariances is provided, the variance-covariance value shown in the title will change accordingly.

When the user pushes the button 'View Results', the results will be displayed in a spreadsheet in EXCEL. In this '(*.xls)' file, the parameter values and optimal designs are provided along with their corresponding labels.

Notice that all plots are by default 'Grid off'. By pushing this button the 'Grid on' property is activated, i.e. grid lines will be added to the current plot. There is also a push button to 'Continue' in the program, i.e. coming back to the main menu. Finally, there is a button to 'Exit' the program.

Figure 6.3(Left) shows that the optimal designs for complete data vary

Figure 6.4: RE 's between optimal designs assuming complete data and data with dropout for clinical trial example.



in function of the serial correlation. If $\rho \leq 0.10$, then the first three measurements should be taken within the first(last) 4 months of the study. If $\rho > 0.10$, the equally-spaced design is very close to the optimal design. (In this example, we have activated the grid option.)

The next step is to compute the design for a specific response probability function. To do that press 'Continue' on the screen DDq_OPTIMAL. The user is send back to the main menu. Here, push down the button 'Data with DROPOUT and HOMOGENEOUS error variance' and select the response probability function 'Other', i.e. the probability function of Table 6.1 will be used. The results are shown in Figure 6.3(Right).

It can be seen that optimal design points for data with dropout are moderately displaced with respect to those for complete data and that the direction of the displacements is leftwards. The results also show that when the autocorrelation parameters is very large $\rho > 0.8$, the follow-up evaluations should be taken in the first half of the follow-up, i.e. when the response probability is the highest.

To compute the relative efficiency of choosing the optimal design(s) for

complete data i.e., Figure 6.3 (Left) instead of those design(s) for data with dropout (same figure but right) is done as follows. Press the button 'There is DROPOUT' situated at the main menu screen and thereafter, select the sort of relative efficiency he/she wants to compute. There are three options in this menu: 'Complete data—linear dropout', 'Complete data—quadratic dropout' and 'Complete data—other dropout'. Choose the third option and press the button 'Plot Results'. The obtained efficiencies are displayed in Figure 6.4. They are very high, i.e. in the range $[0.996, 0.975]$ with ρ being in the range $[0, 0.90]$.

It means that for the example on randomized clinical trial, there is not much efficiency loss if optimal designs for complete data are used instead of those for data with dropout.

Last, the relative efficiencies between the used design $\xi_{\text{example}} = \{-1 \ -0.4596 \ 0.0810 \ 0.5173 \ 1\}$ (after transformation and re-scaling) and its optimal design $\xi_{d,\text{example}}^*$ is provided in Figure 6.5. (This EXCEL spreadsheet is obtained by pressing on the button 'View Result Values'). The *RE*'s are in the range $[0.8901, 0.8245]$ for ρ in the range $[0, 0.80]$. Notice that the *RE* of the correlated cases is moderate, i.e. lower than the reference value 0.85.

Figure 6.5: Spreadsheet with results of relative efficiency of *D*–optimal design between complete data and data with dropout

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
	rho	LB	xopt1	xopt2	xopt3	UB1	d11	d12	d21	d22	p	Prob resp1	Prob resp2	Prob resp3	Prob resp4	Prob resp5	f	RE
1	0.001	-1	-0.6948	0.139307	0.627777	1	0.03	-0.00402	-0.00402	0.005	2	116	88	38	26	27	0.000032	0.996
2	0.101	-1	-0.58165	-0.102	0.367761	1	0.03	-0.00402	-0.00402	0.005	2	116	79	48	31	27	0.00005	0.991
3	0.201	-1	-0.57316	-0.14669	0.278747	1	0.03	-0.00402	-0.00402	0.005	2	116	78	51	33	27	0.000056	0.988
4	0.301	-1	-0.57991	-0.16242	0.214458	1	0.03	-0.00402	-0.00402	0.005	2	116	79	53	35	27	0.000059	0.985
5	0.401	-1	-0.69613	-0.21777	0.16609	1	0.03	-0.00402	-0.00402	0.005	2	116	80	55	37	27	0.000059	0.983
6	0.500	-1	-0.61662	-0.25098	0.12791	1	0.03	-0.00402	-0.00402	0.005	2	116	82	57	38	27	0.000056	0.982
7	0.600	-1	-0.63861	-0.28235	0.096129	1	0.03	-0.00402	-0.00402	0.005	2	116	83	58	40	27	0.000051	0.981
8	0.700	-1	-0.65981	-0.31211	0.067857	1	0.03	-0.00402	-0.00402	0.005	2	116	85	60	41	27	0.000044	0.979
9	0.800	-1	-0.6739	-0.34087	0.040845	1	0.03	-0.00402	-0.00402	0.005	2	116	87	62	42	27	0.000033	0.976
10	0.900	-1	-0.6994	-0.37042	0.01234	1	0.03	-0.00402	-0.00402	0.005	2	116	88	64	43	27	0.000019	0.975
11																		
12																		

6.4.2. Example 2 (Revisited): Longitudinal study of dementia populations

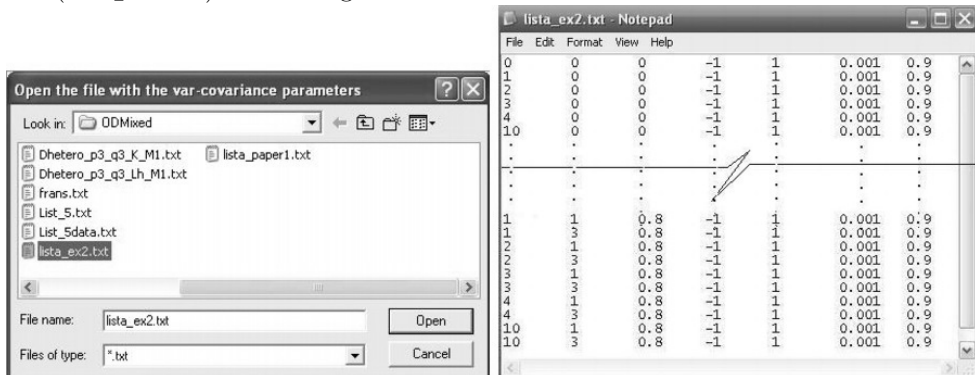
In this example, the decline of dementia patients treated in special care units was studied (Holmes et.al., 2000). The data reported was collected using three equally-spaced design points, i.e at 0, 6 and 12 months, respectively.

In the main screen (Figure 6.1), the number of measurements per subject ($q = 3$), the number of fixed parameters ($p = 2$) and the total number of measurements ($N = 558$ i.e. 186 patients \times 3 time-points) must be filled in.

The best model for the heterogeneity of the error structure is $\Sigma_{jj'} = \sigma^2 \rho_{jj'}^{\text{month}}$, where the error variance and the serial correlation parameter are 0.487 and 0.668, respectively. Although, it was mentioned that the model fit

improves when random-effects are added, no specific random variance structures are provided. So, the check-box panel 'Input variance-covariance data' should be used to provide a list of potential random variance-covariance structures by selecting the option 'Open file with variance-covariances'. The user will then have access to a screen similar to shown in Figure 6.6 (Left) to open the file '(lista_ex2.txt)'. The layout of the input data file will look like in Figure 6.6 (Right).

Figure 6.6: Left: Screen to select '(*.txt)' data file. Right: Layout of the file '(lista_ex2.txt)' containing the variance-covariance data

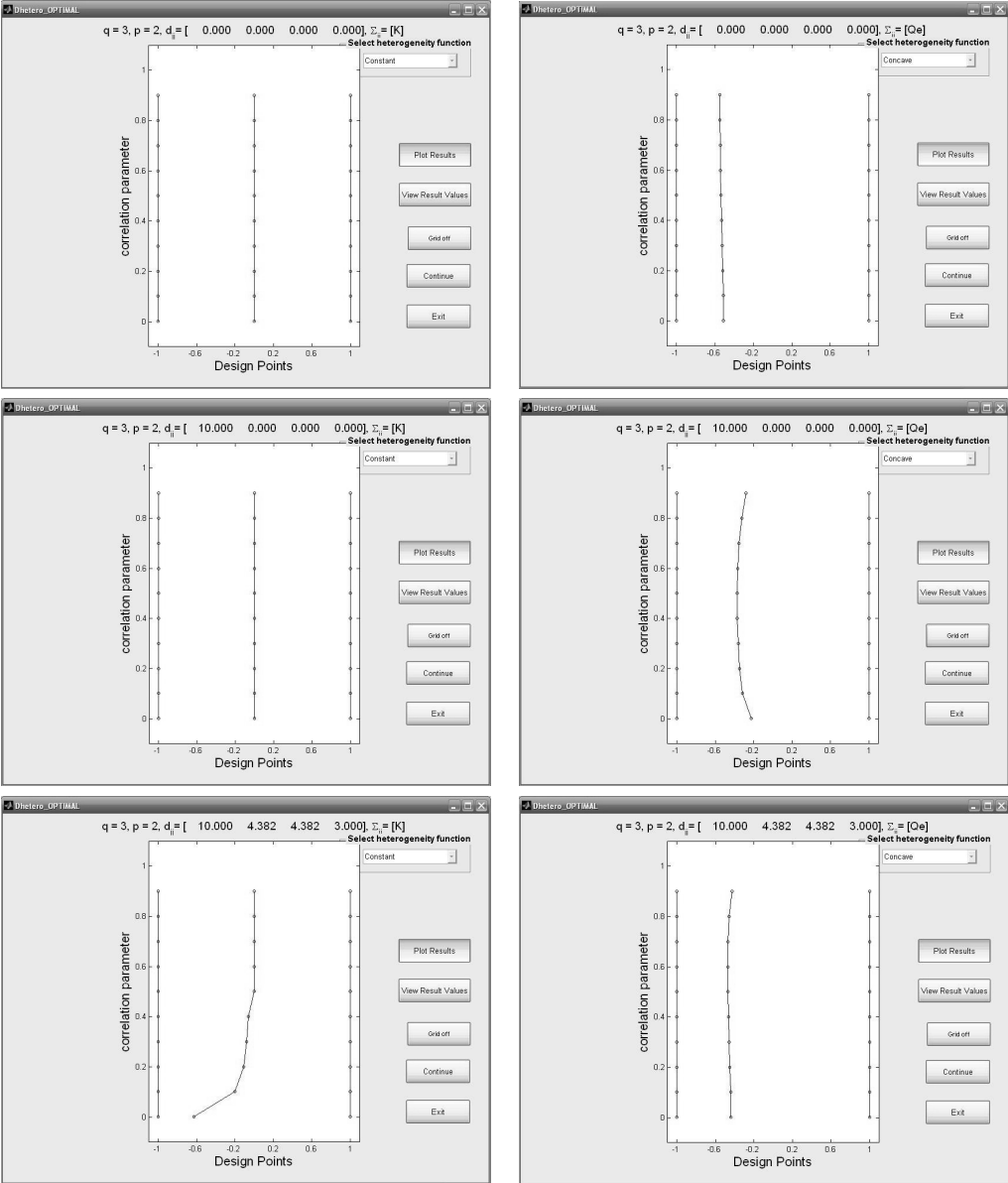


To obtain the optimal designs, the user needs to use the panel 'Compute D -optimal designs when you have' and press the push button 'COMPLETE data with HETEROGENEOUS variance'. Next, he/she will 'select a heterogeneity function', i.e. 'Constant', 'Linear', 'Concave', 'Convex' and 'Other'. A constant heterogeneity function will provide the D -optimal designs for the homogeneous error variance structure (Figure 6.7, Left). By selecting the 'Concave' heterogeneity function for the errors, the D -optimal designs for this heterogeneous error variance structure are obtained. The obtained designs are displayed in Figure 6.7 (Right).

In the upper and middle row plots of Figure 6.7, i.e. when $D = \mathbf{0}$ and $D = [10]$, it can be seen that, when the an homogeneous error structure is assumed, there is no shift of the middle design point as when the error variance structure is considered heterogeneous. The middle point of heterogeneous optimal designs are shifted towards the left of the design space, i.e. where the variances are smaller. Notice that in the lower row plots, i.e. when $D = [10 \ 4.38; 4.38 \ 3]$ the middle points shift still more to the left.

The user can compare the presented design by computing their RE 's. The push button 'There is HETEROSCEDASTICITY' is for this purpose. A screen with the panel menu 'Compute relative efficiency' is displayed.

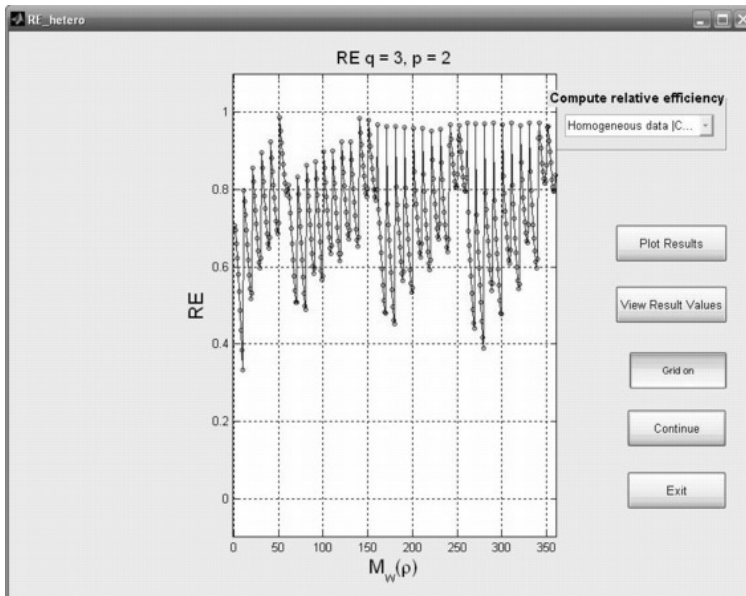
Figure 6.7: Left: D -optimal design assuming error variance is homogeneous.
Right: D -optimal design when error variance structure is heterogeneous.



Using this menu, the user can choose which comparison he is interested on. We have chosen to compare the homogeneous optimal designs with those of a concave function.

By looking at Figure 6.8, it can be seen that the RE 's of the D -optimal designs are strongly connected to the value of the serial correlation. Highly correlated designs have low RE . The random intercept and random slope variance values show a lower effect on the RE 's.

Figure 6.8: RE 's between an homogeneous optimal design and its heterogeneous counterpart for the longitudinal data example.



Notice that using equally-spaced designs $\xi = \{0 \ 6 \ 12\}$ (months), i.e. homogeneous designs for fixed-effects model, instead of using the heterogeneous design $\xi^* = \{0 \ 2 \ 12\}$ (months) is inefficient. The loss of efficiency of the used homogeneous optimal designs with misspecified error variance structure will be in the range $[0.30, 0.65]$. Hence, if ξ^* is applied, the number of patients of special care units will range between $[56, 121]$ which means saving the costs of, at least, 60 patients.

6.5. Discussion and conclusions

In this paper, the use of the ODMIXED program is explained. Optimal designs are found when heterogeneous error structures are expected in the data and(or) when dropout arises during the follow-up of longitudinal stud-

ies. ODMIXED is the first tool tackling dropout and heteroscedasticity of the error and we hope that its user friendly set-up encourages its use for research and teaching purposes. Relative efficiencies in two practical design scenarios are compared to allow the comparison among alternative optimal designs.

A program that computes highly efficient design for a wide range of variance-covariance matrices and another program to compute A -optimal designs for linear mixed models are also developed. The layout of the graphical user interphase of these two programs is similar to that of ODMIXED. The former program is based on Ortega, Tan and Berger (2009), while the latter is under revision. These programs will appear elsewhere.

Future research is encouraged to extend this program to the case when the collected data have heterogeneous error as well as dropout.

Appendix: Created files

In Table 6.2, we summarize the list of files created by the functions associated to each push button of the third and fourth menus of the graphical user interphase. Depending on the push button, either two or three sorts of files are created during the execution of the program ODMIXED. The files with '*.txt' format can be open in any wordprocessor (NOTEPAD, WORD); the files with '*.xls' format can be open in EXCEL and the files with '*.fig' have to be open in MATLAB'.

Table 6.2: Files created by ODMIXED.

Name push button	Name of file created	Format	Description
'COMPLETE data with HETEROGENEOUS error variance'	Dhetero_p3_q3_K_M1.txt	*.txt	Optimal designs for $p = 3$ $q = 3$; 'K','Lh','Qe','Qq' correspond to constant, li- near, concave and concave functions; (M1) refers to the first var-covar.struct. File *_ALL_K.* has the re- sults of $M1 \dots Mn$ var-cov.str.
	Dhetero_p3_q3_Lh_M1.txt	*.xls	
	Dhetero_p3_q3_ALL_K.txt		
	Dhetero_p3_q3_ALL_Lh.txt		
	Dhetero_p3_q3_ALL_K.xls		
	Dhetero_p3_q3_ALL_Lh.xls		
'Data with DROPOUT and HOMOGENEOUS error variance'	Ddropout_p3_q3_F_M1.txt	*.txt	Optimal designs for $p = 3$ $q = 3$; 'K','L','Q','O' refer to constant, linear, quadratic and other functions for the M1 var-cov. struct. File *_ALL_F.* has the re- sults of $M1 \dots Mn$ var-cov.str.
	Ddropout_p3_q3_L_M1.txt	*.xls	
	Ddropout_p3_q3_ALL_F.txt		
	Ddropout_p3_q3_ALL_L.txt		
	Ddropout_p3_q3_ALL_F.xls		
	Ddropout_p3_q3_ALL_L.xls		
'COMPLETE data with HOMOGENEOUS error variance'	DDq_p2_q5_M1.txt	*.txt	Optimal designs for $p = 2$ $q = 5$; M1 or ALL stand for the var.conv. structure. for the first or ALL model(s) File *.fig contains the RE plot. (TB') File contains table in EXCEL.
	DDq_p2_q5_M1.fig	*.xls	
	DDq_p2_q5_ALL.txt	*.fig	
	DDq_p2_q5_TB.xls		
#fixed parameters (p) < # measurements (q) 'There is HETEROS- CEDASTICITY' 'There is DROPOUT'	RE_p2_q3.fig	*.xls	Relative efficiency for varying the value of $p = 2$ given at fixed $q = 3$. RE homogeneous w.r.t. (L') linear variance func. RE complete(Full) data w.r.t. linear dropout. File *.fig contains the RE plot. (TB') File contains table in EXCEL.
	RE_p2_q3_TB.xls	*.fig	
	RE_hetero_p2_q3_L.xls		
	RE_dropout_p2_q3_F_L.xls		
	RE_dropout_p2_q3_F_L.fig		

References

- Abt, M., Liski, E.P., Mandal, N.K., and Sinha, B.K.(1997). Optimal designs in growth curve models: Part I correlated model for linear growth: optimal designs for slope parameter estimation and growth prediction. *Journal of Statistical Planning and Inference*, **64**, 141–150.
- Abt, M., Gaffke, N., Liski, E.P., and Sinha, B.K.(1998). Optimal designs in growth curve models: Part II correlated model for quadratic growth: optimal designs for slope parameter estimation and growth prediction. *Journal of Statistical Planning and Inference*, **67**, 287–296.
- Antille, G. and Weinberg, A.(2000). *A study of D-optimal efficiency for polynomial regression*. Université de Genève. Cahiers du département d'économétrie, Faculté des sciences économiques et sociales. Report No.2000.04. Geneve, Switzerland. <http://www.unige.ch/ses/metri>.
- Atkins,J.E., and Cheng, C.S.(1999). Optimal regression designs in the presence of random block effects. *Journal of Statistical Planning and Inference*, **77**, 321–335.
- Atkinson, A.C. and Cook, R.D.(1995). D-optimum Designs for heteroscedastic linear models. *Journal of the American Statistical Association*, **90**(429), 204–212.
- Atkinson, A.C., and Donev, A.N.(1996). *Optimum Experimental Designs*. Oxford: Clarendon Press.
- Berger, M.P.F., Ouwens, M.J.N.M., and Tan, F.E.S.(2002). Robust designs for longitudinal mixed effects models. In H.Yanai et.al., editor, *Advances in Psychometric Research*, 339–346. Springer Verlag.

- Berger, M.P.F., and Tan, F.E.S.(2004). Robust designs for linear mixed effects models. *Journal of the Royal Statistical Society*, series C, **53**(4), 569–581.
- Bischoff, W. (1993), On D -optimal designs for linear models under correlated observations with an application to a linear model with multiple response. *Journal of Statistical Planning and Inference*, **37**, 69–80.
- Bischoff, W. (1996). On maximin designs for correlated observations. *Statistics and Probability Letters*, **26**, 357–363.
- Bouter, L.M., van Dongen, M.C.J.M, and Zielhuis, G.A.(2005). *Epidemiologisch onderzoek: Opzet en interpretatie*. Houten: Bohn Stafleu van Loghum.
- Bunke, H., and Bunke, O.(1986). *Statistical Inference in Linear Models* vol. I. Berlin: John Wiley & Sons Ltd.
- Carroll, K.M., Rounsaville, B.J., Nich, C., Gordon, L.T., Wirtz, P.W., and Gawin, F.(1994). One-year follow-up of psychotherapy and pharmacotherapy for cocaine dependence: Delayed emergence of psychotherapy effects. *Archives General Psychiatry*, **51**, 989–997.
- Chaloner, K., and Verdinelli, I.(1995). Bayesian experimental design: a review. *Statistical Science*, **10**, 273–304.
- Cheng, C.S.(1999), Optimal regression designs under random block-effects models. *Statistica Sinica*, **5**, 485–497.
- Conlinsk J., and Watts H. (1979), A model for optimizing experimental designs for estimating response surfaces. *Journal of Econometrics*, **11**, 27–42.
- Crisp, A. and Burrige, J.(1994). A note on nonregular likelihood functions in heteroscedastic regression models. *Biometrika*, **81**, 585–587.
- Dette, H. (1997). Designing experiments with respect to standardized optimality criteria. *Journal Royal Statistical Society B*, **59**(1), 97–110.
- Dette, H., Martinez,I., Ortiz, I.M., and Pepelyshev, A.(2006). Maximin efficient design of experiment for exponential regression models. *Journal of Statistical Planning and Inference*, **136**(12), 4397–4418.
- Dette, H., and Neugebauer, H.M.(1997). Bayesian d-optimal designs for exponential regression models. *Journal of Statistical Planning and Inference*, **60**(2), 331–349.
- Dette H., and O'Brien T.E.(1999), Optimality criteria for regression models based on predicted variance. *Biometrika*, **86**, 93–106.

- Dette, H., Song, D., and Wong, W.K.(2001). Robustness properties of minimally supported Bayesian D-optimal designs for heteroscedastic models. *The Canadian Journal of Statistics*, **29**, 633–647.
- Diggle, P.J., Liang K., and Zeger, S.L. (1994). *Analysis of Longitudinal Data*. Oxford: Oxford University Press.
- Elving, G.(1959) Design of linear experiments. In *Probability and Statistics:The Harald Cramér Volume (U. Grenander, ed.)* 58–74, Wiley, New York.
- Everitt B.S. (1998) Analysis of longitudinal data: Beyond MANOVA. *British Journal of Psychiatry*. (Review article), **172**, 7–10.
- Fang, Z., Wiens, D.P., and Wu, Z.(2006). Locally D-optimal designs for multi-stage models and heteroscedastic polynomial regression models. *Journal of Statistical Planning and Inference*, **136**, 4059–4070.
- Fedorov, V.V. (1972). *Theory of Optimal Experiments*. New York: Academic Press.
- Fukushima, M., Colmerauer, M.E.M., Nayak, S.K., Koziol, J.A., and Pilch, Y.H. (1982a) Immunotherapy of a murine colon cancer with syngeneic spleen cells, immune RNA and tumor antigen. *International Journal of Cancer*, **29**, 107–112.
- Fukushima M., Colmerauer, M.E.M., Nayak, S.K., Koziol, J.A., and Pilch, Y.H.(1982b) Antitumor effect of syngeneic spleen cells treated with immune RNA and tumor antigen. *International Journal of Cancer*, **29**, 113–117.
- Galbraith, S. and Marschner, I.C.(2002). Guidelines for the design of clinical trials with longitudinal outcomes. *Controlled Clinical Trials*, **23**, 257–273.
- Gram, L.F., Reisby, N., Ibsen, I., Nagy, A., Dencker, S.J., Bech,P., Petersen, G.O., and Christiansen, J. (1976). Plasma levels and antidepressive effects of imipramine. *Clinical Pharmacological Therapy*, **19**, 318–324.
- Goldstein, H. (1976). *The Design and analysis of Longitudinal Studies: Their Role in the Measurement of Change*. London: Academic Press, London.
- Hackl, P.(1995). Optimal design for experiments with potentially failing trials. Kitsos, C.P. and Müller, W.G. In *Proceedings of MODA4: Advances in Model-Oriented Data Analysis*, Physica Verlag,Heidelberg, 117–124.
- Hedeker, D., Gibbons, R.D. Waternaux, C. and Davis, J.M. (1989). Investigating drug plasma levels and clinical response using random regression models. *Psychopharmacology Bulletin*, **25**, 227–231.

- Hedeker, D. and Gibbons, R.D.(1996). Mixreg: a computer program for mixed-effects regression analysis with autocorrelated errors. *Computer Methods and Programs in Biomedicine*, **49**(3), 229-252.
- Herzberg, A.M. and Andrews, D.F.(1976). The robust design of experiments. *Bulletin of the International Statistical Institute*, **46**(3), 370-374.
- Herzberg, A.M. and Andrews, D.F.(1976). Some considerations in the optimal design of experiments in non-optimal situations. *Journal of the Royal Statistical Society, Serie B*, **38**(3), 284-289.
- Holmes, D., Teresi, J., and Kong, J.(2000) Service inputs and costs of care related to outcomes among cognitively impaired nursing home residents. *Journal of Mental Health Policy and Economics*, **3**, 121-127.
- Huber, P.J. (2004). *Robust statistics*. Cambridge: John Wiley & Sons.
- Huele, A.F. (1998). *Statistical Robust Design*. PhD Dissertation, University of Amsterdam.
- Imhof, L.A., Song, D., and Wong, W.K.(2002). Optimal design of experiments with possibly failing trials. *Statistica Sinica*, **12**, 1145-1155.
- Imhof, L.A., Song, D., and Wong, W.K.(2004). Optimal designs of experiments with anticipated pattern of missing observations. *Journal of Theoretical Biology*, **228**, 251-260.
- Ishak, K.J., Platt, R.W., Joseph, L., Hanley, J.A., and Caro, J.J. (2007). Meta-analysis of longitudinal studies. *Clinical Trials*, **4**, 525-539.
- Keller, A., Castellanos, F.X., Vaituzis, A.C., Jeffries, N.O., Giedd, J.N. and Rapoport, J.L. (2003). Progressive loss of cerebellar volume in childhood-onset schizophrenia. *The American Journal of Psychiatry*, **160**, 128-133.
- Kiefer, J. (1959). Optimal experimental designs (with discussion). *Journal of the Royal Statistical Society B*, **21**, 272-319.
- Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. *Canadian Journal of Mathematics*, **12**, 363-366.
- Kiefer, J. and Wolfowitz, J.(1961). Optimum designs in regression problems. *Annals of Mathematical Statistics*, **30**, 271-294.
- Kiefer, J.(1974). General equivalence theory for optimum designs (approximate theory). *The Annals of Statistics*, **2**(5), 849-879.
- Kiefer J.(1975). Optimal design: Variation in structure and performance under change of criterion. *Biometrika*, **62**(2), 277-288.

- Lagarias, J.C., Reeds, J.A., Wright, M.H., and Wright, P.E. (1998). Convergence properties of the nelder-mead simplex method in low dimensions. *SIAM Journal of Optimization*, **9**, 112-147.
- Laird, N.M., and Ware, J.H. (1982). Random-effects models for longitudinal data. *Biometrics*, **38**, 963-974.
- Little, R.J.A., and Rubin, D.B. (1987). *Statistical Analysis with Missing Data*. New York: Wiley.
- MATLAB (2004), *Handbook MATLAB, version 7.0.1. (R14)*. MathWorks Inc, MA, USA.
- Mentré F., Mallet A., and Baccar D.(2006). Optimal designs in random effects regression models. *Biometrika*, **84**, 429-442.
- Moen, R.D., Nela T.W., and Provost L.P. (1991). Improving Quality throughout Planned Experiments.
- Moerbeek, M. (2004), Robustness properties of A-, D-, and E-optimal designs for polynomial growth models with autocorrelated errors. *Computational Statistics and Data Analysis*, **48**, 765-778.
- Molenberghs, G., and Kenward, M.G. (2007). *Missing Data in Clinical Studies*. London: John Wiley & Sons.
- Molenberghs, G., Tuerlinckx, F., and Verbeke, G.(2006), *Numerical Techniques for Statisticians*, Universiteit Hasselt and Katholiek Universiteit Leuven.
- Montepiedra, G. and Wong, W.K.(2001). A new design criterion when heteroscedasticity is ignored. *Annals of the Institute of Statistical Mathematics*, **53**, 418-426.
- Montgomery, P.C.(2000). *Design and Analysis of Experiments* (5th ed.). New York: John Wiley & Sons.
- Müller, Ch. H. (1995). Maximin efficient designs for estimating nonlinear aspects in linear models. *Journal of Statistical Planning and Inference*, **44**, 117-132.
- Näther, W. (1985). *Effective Observation of Random Fields*, Leipzig: Teubner-Texte.
- Nich, C., and Carroll, K.(1997). Now you see it, now you don't. A comparison of traditional versus random-effect regression models in the analysis of longitudinal follow-up data from a clinical trial.*Journal of Consulting and Clinical Psychology*, **65**(2), 252-261.

- Ortega, S.A., Tan, F.E.S., and Berger, M.P.F.(2008). The effect of dropout on the efficiency of D-optimal designs of linear mixed models. *Statistics in Medicine*, **27**(14), 2601–2617.
- Ortega, S.A., Tan, F.E.S., and Berger, M.P.F.(2009). Highly efficient designs to handle the incorrect specification of linear mixed models. *Communications in Statistics - Simulation and Computation*, **38**(1), 14–30.
- Ouwens, M.J.N.M., Tan, F.E.S. and Berger, M.P.F.(2002), Maximin D-optimal designs for longitudinal mixed effects models. *Biometrics*, **58**, 735–741.
- Ouwens, M.J.N.M., Tan, F.E.S. and Berger, M.P.F.(2006), A maximin criterion for the logistic random intercept model with covariates. *Journal of Statistical Planning and Inference*, **136**(3), 962–981.
- Pearson, K. (1895). Contributions to the mathematical theory of evolution, II. Skew variation in homogeneous material. *Phil. Transactions Royal Society of London*, **186**, 343–414.
- Petkova, E. and Teresi, J.(2002). Some statistical issues in the analyses of data from longitudinal studies of elderly chronic care populations. *Psychosomatic Medicine*, **64**, 531–547.
- Polit, D.F., and Beck, C.T. (2004). *Nursing Research: Principles and Methods* (7th ed.). Philadelphia: Lippincott Williams & Wilkins.
- Pukelsheim, F. (1993). *Optimal Designs of Experiments*. New York: John Wiley & Sons.
- Rao, C.R.(1973).*Linear Statistical Inference and Its Applications*, New York: Wiley.
- Rao, R.S., Prakasham, R.S., and Hobbs, P.S. (2008) The Taguchi methodology as an statistical tool for biotechnological applications: a critical appraisal. *Biotechnology Journal*, **3**(4), 510–523.
- Reisby, N. Gram, L.F., Bech, P., Nagy, A., Petersen, G.O., Ortmann, J., Ibsen, I., Dencker, S.J., Jacobsen, O., Krautwald, O., Sondergaard, I., and Christiansen J. (1977). Imipramine: Clinical effects and pharmacokinetic variability. *Psychopharmacology*, **54**, 263–272.
- Rosenbaum, P. (2002). *Observational Studies* (2nd ed.). New York: Springer.
- Scheffé, H. (1959). *The Analysis of Variance*. New York: Wiley.
- Snedecor, G. (1937). *Statistical Methods*. Iowa: The Iowa State College Press, Ames.

- Taguchi G., and Konishi S., (1987). *Orthogonal Arrays and Linear Graphs*. Dearborn, MI: ASI Press.
- Tan, F.E.S., and Berger, M.P.F. (1999) Optimal allocation of time points for the random effects models. *Communications in Statistics, Simulations and Computations*, **28**(2), 517–540.
- Taris, T.W. (2000). *A Primer in Longitudinal Data Analysis*. London: Sage Publications.
- Tekle, F.B., Tan, F.E.S., and Berger, M.P.F.(2008a). D-optimal cohort designs for linear mixed-effects models. *Statistics in Medicine*, **27**(14), 2586–2600.
- Tekle, F.B., Tan, F.E.S., and Berger, M.P.F.(2008b). Highly Efficient Designs for Logistic Models with Categorical Variables. *Communications in Statistics - Theory and Methods*, **37**(5), 746–759.
- Tekle, F.B., Tan, F.E.S., and Berger, M.P.F.(2009). Interactive computer program for optimal designs of longitudinal cohort studies. In press *Computer Methods and Programs in Biomedicine*, DOI:10.1016/j.cmpb.2008.11.002.
- Tsai, K.T., and Koziol, J.A. (1988). VARCOV II: a computer program to multivariate analysis of growth and response curves. *Computer Methods and Programs in Biomedicine*, **27**, 69–74.
- Tukey, J.W. (1960). A survey of sampling from contaminated distributions. In I. Olkin, *Contributions to Probability and Statistics*. Stanford: Stanford University Press.
- Verbeke, G., and Lesaffre, E. (1999). The effect of drop-out on the efficiency of longitudinal experiments. *Applied Statistics*, **48**(3), 363–375.
- Verbeke, G., and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer.
- Wong, W.K. (1994). Comparing robust properties of A, D, E and G-optimal designs. *Computational Statistics and Data Analysis*, **18**, 441–448.
- Wong, W.K.(1995). On the equivalence of D - and G - optimal designs in heteroscedastic models. *Statistics & Probability Letters*, **25**, 317–321.
- Wong, W.K. and Cook, R.D.(1993). Heteroscedastic G-optimal designs. *Journal of the Royal Statistical Society, Series B*, **4**, 871–880.
- Wynn, H.P. (1970). The sequential generation of D-optimum experimental designs. *The Annals of Mathematical Statistics*, **41**(5), 1655–1664.

Summary and recommendations

In this dissertation, the focuss is on the robustness of designs against the incorrect specification of the model, the covariance structure assumptions and the design itself. Below a summary of each chapter is provided. Thereafter, a set of practical guidelines is put forward for researchers dealing with longitudinal studies. Last, some suggestions for future research are made.

Chapter 1 has four purposes: (1) to explain the need of robust (longitudinal) designs; (2) to introduce the theoretical background underlying the finding of robust designs; (3) to explain the scope of the research; and (4) to introduce the organization of this dissertation. The theoretical background includes the formulation of the linear mixed model and its assumptions, the concept of response probability, the ideas behind the D - and A - optimality criteria and the way to identify robust designs. By means of a clinical trial example on cancer growth in mice, connections between the model assumptions and the concept of robustness are put forward.

Chapter 2 is devoted to study designs that handle the incorrect specification of the model parameters and of the between-subjects variance structure. A maximin criterion is applied to examine the relative efficiency of several optimal designs for a family of linear mixed models. The main conclusions are (1) the maximin design encountered appears to be highly efficient; (2) the variation of the minimum relative efficiencies corresponding to the optimal designs of the family of linear mixed models that were studied decreases as the order of the polynomial increases; (3) the effect of the autocorrelation parameter on the relative efficiencies of optimal designs

is the largest for first-degree polynomials; and (4) the relative efficiency of the equidistant design is lower than that of the maximin value. To illustrate the results, a longitudinal study on an antidepressive drug treatment for depressive inpatients, as reported by Gram et.al. (1976) and Reisby et.al (1977), is employed.

Chapter 3 focuses on the incorrect specification of within-subjects variance structure. The loss of efficiency of designs which assume an homogeneous error variance structure (instead of the heterogeneous one) is investigated. The main conclusions are (1) the optimal designs of heterogeneous linear mixed models show moderate-to-large shifts in locations of the time points as compared with those having an homogeneous error variance structure; (2) if the error variance structure follows a straight-line model, the optimal designs for homogeneous data remain highly efficient regardless the number of parameters, the degree of the polynomial, the serial correlation and the random variances and covariances; and (3) efficiencies are not always high when the error variance follows a convex or a concave trajectory. As example a decade-long longitudinal study to model the decline of mental state (dementia) of elderly chronic care populations, as published by Holmes et.al. (2000) and Petkova and Teresi (2002), is used.

Chapter 4 concerns D -optimal designs for linear mixed models when dropout is encountered. The efficiency loss is estimated in cases where a D -optimal design for complete data is chosen instead of that for data with dropout. Two types of monotonically decreasing response probability functions are investigated to describe dropout. The results show that (1) for the dropout case the locations of the D -optimal design points will shift with respect to those for the complete and uncorrelated data case; (2) due to this shift, the information collected at the D -optimal design points for the complete data case does not correspond to the smallest variance; and (3) the size of the displacement of the time points depends on the linear mixed model and the efficiency loss is moderate. To illustrate the results, data from a 1-year follow-up of a randomized clinical trial of psychotherapy and pharmacotherapy for ambulatory cocaine addicts reported by Carroll et.al. (1994) and Nich et.al. (1997), is considered.

Chapter 5 deals with the robustness of A -optimal designs for (serially) correlated random-intercept models with respect to their design scale and connected weights. The relative efficiency measures the loss of efficiency when designs that are optimal in one scale are applied to another scale. The results show that (1) small to moderate efficiency losses arise when designs that are optimal in a specific narrow scales, e.g., $[-1, 1]$, $[0, 1]$, or $[0, 2]$, are applied on another narrow scale; (2) moderate to large efficiency losses arise

when designs which are optimal in a narrow scale are applied to a broad scale, e.g., $[0, 10]$, and vice-versa; (3) the performance of equally-weighted A -optimal designs is not good when compared to their connected designs with unequal weights; and (4) the designs which are optimal in the scale $[0, 10]$ are the least efficient.

Chapter 6 introduces a tool for researchers. ODMIXED is a computer program to obtain optimal designs for linear mixed models. Designs that account for heterogeneous-correlated errors and for data with dropout are obtained and compared by using relative efficiencies. Two examples (Chapter 3 and 4) are worked out to illustrate the application of optimal design theory at the planning stage of longitudinal studies.

7.1. Guidelines for researchers

This dissertation has as a spin-off some general and specific practical guidelines. These guidelines aim at simplifying the researcher's task of finding and selecting optimal/robust designs for longitudinal studies when assumptions on the model and/or its variance-covariance structure are (partially) misspecified.

The general guidelines are

1. Researchers should use prior information on the underlying model to put forward as explicitly as possible their model assumptions before choosing a design to collect longitudinal data. This is because the number of fixed parameters p and the presence/absence of serial correlation have a larger effect on the optimal design and corresponding relative efficiency than the values of the parameters do.
2. Researchers should be aware that it is generally not very efficient if the number of repeated measurements exceeds the order of the polynomials plus one. Increasing the number of repeated measurements does not lead to much more information on the values of the underlying model parameters.

The specific guidelines for polynomials of at most order three depend on the amount of information one has about the missingness mechanism and about the underlying model. These are summarized as follows:

1. Suppose that the longitudinal design can be assumed to be complete.
 - a. The researcher may select the four-points equally-spaced design as a conservative option. If no additional information about the

serial correlation ρ is available, the equally spaced designs has a relative efficiency of $RE = 0.84$ (see Table 2.4). The relative efficiency becomes as large as $RE = 0.92$ if it is known that there is serial correlation.

- b. Suppose that the error variances are homogeneous. If no additional information about the serial correlation ρ is available, then any of the maximin designs provided in Table 2.2 of Chapter 2 can be chosen. The corresponding relative efficiencies remain above 0.87. If it is known that there is serial correlation, the maximin design in Table 2.3, i.e., $\xi_{MMV} = \{-1 \quad -0.4247 \quad 0.4247 \quad 1\}$ should be selected. The corresponding relative efficiency is equal to 0.97.
 - c. If the assumption of homogeneous error variance is not tenable, then the designs in Table 2.2 may have low relative efficiencies (efficiency loss up to 62%, Chapter 3). If additional information about how the error variances change with time is available, then much higher efficiencies can be obtained. For example, if the order p of the polynomial is known, the optimal designs as indicated in Table 2.2 have a relative efficiency larger than $RE = 0.91$, provided that the number of repeated measurements is equal to the order of the polynomial plus one.
2. Suppose there is dropout and the underlying missingness mechanism is MAR. If the missingness mechanism is MCAR and independent of the time points, then the optimal design will obviously not change compared to the complete data situation. Due to the MAR dropout the optimal designs are shifted to the left and using designs that are optimal for this specific MCAR case will lead to efficiency loss. To compensate this loss, 15% extra measurements should be added (Chapter 4).
 3. The researcher should prefer D - over A -optimality criterion. If a researcher is uncertain about the suitable time interval for his/her study, A -optimal designs are not recommended because they are not scale invariant, i.e. when there is a large difference between a time interval $[a, b]$ and another time interval $[c, d]$, the designs that are optimal in $[a, b]$ might become very inefficient in $[c, d]$ (Chapter 5).

7.2. Further research

This dissertation has shown that optimal designs are in general quite robust with respect to their model assumptions. However, there is room for improvement. The search of D -robust designs for longitudinal studies is encouraged in the following directions.

1. Find robust designs for polynomial models allowing time varying covariates. In many practical situations, covariates are involved for which the distribution is not determined prior to data collection. The study could be concentrated on the variates for which the joint marginal distribution can be determined prior to data sampling.
2. Find robust designs for linear mixed models when the dropout is Missing Not At Random (MNAR). In practice it is often difficult to justify the assumption of Missing At Random. It would be interesting to fit a selected number of plausible models that allow for MNAR and study the effect on the D -optimal designs.
3. Find robust designs for discrete longitudinal responses. It is known that the D -optimal designs for non-normal models may differ from those for the normal counterpart. It is, however, not known whether robust designs under linear mixed models remain robust under logistic mixed models or GEE models.
4. Find robust designs for linear mixed models under other optimality criteria. For example, when researchers are only interested in a subset of the model parameters, the criteria may affect robustness.

Samenvatting in het Nederlands (Dutch Summary)

Het doel van dit proefschrift is het onderzoeken van de robuustheid van onderzoeksontwerpen of "Research designs" tegen het verkeerd specificeren van het model, de covariantiestructuur veronderstellingen en het ontwerp zelf. Hieronder wordt een samenvatting van elk hoofdstuk gegeven. Vervolgens worden een aantal praktische richtlijnen gegeven voor onderzoekers die met longitudinale studies werken. Tenslotte worden enkele aanbevelingen voor toekomstige onderzoek gegeven.

Hoofdstuk 1 heeft vier doelen: (1) het verklaren van de behoefte aan robuuste (longitudinale) ontwerpen; (2) het introduceren van de theoretische achtergrond die ten grondslag ligt aan het vinden van robuuste ontwerpen; (3) het uitleggen van het onderzoekskader; en (4) het introduceren van de opzet van het proefschrift. De theoretische achtergrond omvat ook de formulering van het lineaire mixed-effects model en haar aannames, het responswaarschijnlijkheidsconcept, de ideeën D - en A -optimaliteitscriteria en de methoden om robuuste ontwerpen te vinden. Door middel van een klinische experiment over de groei van kanker cellen bij muizen, worden relaties tussen de modelaannames en het robuustheidsconcept gelegd.

Hoofdstuk 2 is gewijd aan het onderzoeken van robuuste ontwerpen tegen het verkeerd specificeren van de modelparameters en de tussen-persoonsvariantie structuur. Een maximin-criterium wordt toegepast om de relatieve efficiëntie van verschillende optimale ontwerpen voor een familie van lineaire mixed-effects modellen te onderzoeken. De belangrijkste conclusies zijn: (1) het gevonden maximin ontwerp blijkt hoog efficiënt te zijn; (2) de variatie

van de minimum relatieve efficiënties behorend bij de optimale ontwerpen van de onderzochte familie van lineair mixed-effects modellen neemt af naarmate de graad van de polynoom toeneemt; (3) de invloed van de autocorrelatie parameter op de relatieve efficiënties van optimale ontwerpen is het grootste voor eerstegraads polynomen; en (4) de relatieve efficiëntie van het equidistante ontwerp is lager dan die van het maximin ontwerp. De resultaten worden verhelderd met een longitudinale studie over het gebruik van een antidepressivum voor depressieve patiënten, zoals beschreven door Gram et.al. (1976) en Reisby et.al (1977).

Hoofdstuk 3 richt zich op de verkeerde specificatie van de binnenpersoonsvariantie structuur. Het efficiëntieverlies van ontwerpen die homogeniteit veronderstellen van de foutenvariantie structuur ten opzichte van ontwerpen die dit niet veronderstellen wordt onderzocht. De belangrijkste conclusies zijn: (1) de optimale ontwerpen van heterogene lineaire mixed-effects modellen vertonen matige tot grote verschuivingen in de locaties van de tijdstippen ten opzichte van ontwerpen met een homogene foutenvariantie structuur; (2) als de foutenvariantie structuur een lineair verloop heeft, dan blijven de optimale ontwerpen voor homogene data hoog efficiënt ongeacht het aantal parameters, de graad van de polynoom, de seriële correlatie en de random varianties en covarianties; en (3) efficiënties zijn niet altijd hoog als de foutenvariantie een convex of concaaf verloop heeft. Als voorbeeld wordt een longitudinale studie gebruikt betreffende de achteruitgang van de geestelijk toestand (dementie) van bejaarde chronische-zorg patiënten van Holmes et.al. (2000) en Petkova and Teresi (2002).

Hoofdstuk 4 betreft D -optimale ontwerpen voor lineaire mixed-effects modellen met uitval(dropout). Het efficiëntieverlies wordt geschat als een D -optimaal ontwerp voor complete data wordt gekozen in plaats van een ontwerp voor data met uitval. Twee soorten van monotoon dalende respons waarschijnlijkheidsfuncties worden onderzocht om uitval te beschrijven. De resultaten tonen aan dat (1) de posities van de D -optimale tijdstippen verschuiven ten opzichte van de posities voor compleet en ongecorrleerde data; (2) door deze verschuiving is de informatie op de D -optimale tijdstippen voor het geval van complete data niet meer de kleinste variantie; en (3) de grootte van de verschuiving van de tijdstippen hangt af van het lineair mixed-effects model en het efficiëntieverlies is matig. Om de resultaten te illustreren worden data gebruikt uit een 1-jarig gerandomiseerd klinisch experiment met psychotherapie en farmacotherapie voor ambulante cocaineverslaafden zoals beschreven door Carroll et.al. (1994) en Nich et.al. (1997).

Hoofdstuk 5 behandelt de robuustheid van A -optimale ontwerpen voor (seriële) gecorrleerde random-intercept modellen, met betrekking tot

de ontwerp schaal en de gewichten. De relatieve efficiëntie meet het efficiëntieverlies als ontwerpen die optimaal zijn in de ene schaal worden toegepast in een andere schaal. De resultaten tonen aan dat (1) lage tot matige efficiëntieverliezen zich voordoen als ontwerpen die optimaal zijn op een specifieke schaal met kleine range, e.g., $[-1, 1]$, $[0, 1]$ of $[0, 2]$, worden toegepast op een andere schaal met kleine range; (2) matige tot grote efficiëntieverliezen doen zich voor als ontwerpen die optimaal zijn in een schaal met kleine range worden toegepast in een schaal met grote range, e.g., $[0, 10]$, en vice versa; (3) A -optimale ontwerpen met gelijke gewichten werken niet goed in vergelijking met de corresponderende ontwerpen met ongelijke gewichten; en (4) de ontwerpen die optimaal zijn in de schaal $[0, 10]$ zijn het minst efficiënt.

Hoofdstuk 6 introduceert een tool voor onderzoekers. ODMIXED is een computerprogramma om optimale ontwerpen voor lineaire mixed models te berekenen. Ontwerpen waarbij rekening wordt gehouden met heterogeen-gecorrleerde fouten en data met uitval kunnen berekend worden en vergeleken worden op basis van hun relatieve efficiënties. Twee voorbeelden (Hoofdstuk 3 en 4) zijn uitgewerkt om de toepassing van optimale ontwerptheorie te illustreren in de planningsfase van longitudinale studies.

Acknowledgements

I am indebted to several people. First of all, I would like to mention dr. Frans Tan and prof. dr. Martijn Berger. Frans and Martijn, thank you very much for your supervision, patience, guidance, and knowledge. I really appreciated the supervision meetings with criticism, motivation, and good mood. Your interest in me as PhD student and as person encouraged me a lot. Frans, thank you for the many suggestions and time you dedicated to me all the way through. Martijn, thank you for your confidence in me, the meetings at NIAS in Wassenaar, and your assertive evaluations. I owe you both so much.

I would like to express my gratitude to prof. dr. M.H. Prins, prof. dr. ir. R.L.M. Peeters, dr. ir. M. Moerbeek, and prof. dr. G. Molenberghs for their comments, remarks, and advices on the first version of my thesis. Warm thanks also go to dr. Gerard van Breukelen for his detailed questions and suggestions. I want to acknowledge the time the evaluation committee spent on reviewing my thesis.

I would like to thank the members and many former members of the group Methodology and Statistics at the University of Maastricht. I learned that teaching statistics is not an easy task and that there is a life next to doing a PhD. Thanks for your friendship and many worthwhile comments. In particular, I would like to thank Fetene Bekele, Micky Chenault, Marion de Leeuw, Bärbel Maus, Stephanie van Nispen, Valeria Lima Passos, Pauline Vilters, Monique Reusken and last but not least Marga Doyle. I will always remember the many encouraging words and good advices you provided me. Thanks for the nice atmosphere.

Many thanks go also to my previous supervisors in Eindhoven, Stef van Eijndhoven, Jan B. Dijkstra, Jan Engel, Freek Huele, and Emiel van

Berkum. Thanks a lot for your advice and support during my time at the TU/e and for giving me the opportunity to work with semi-parametric mixed models in an applied context.

I want to acknowledge the prompt and effective help of Gert-Jan Meeuwis, Anne Schreiber, and André Verwij last year. Thanks for not keeping a fence between work and personal issues but seeking the best mix between them.

The last years would not have been possible without the support of friends. I would like to thank Els, Martine, Victor, Jeroen, Elmar, Friso, Teis, Marloes, Ivo, Govert, Eline, Frans, Frank, Jan, Mihaela, Rik, Eunice, Vincent, Marcha, Tasmara, Arjan, Stefania, Michel, Frans, Carola, Andrea, Bérénice, Marien, Ingrid, Lena, Ioana, and Jan Willem. Many thanks go also to Olga Maldonado, Sabine Schnabel, Eleni Sarantopoulou, Anton Tijhuis, Magaly Valdivia, Rosemary and Albertico Gumbs-Lopez, and Sra. Benigna Canaza. I cannot tell how much I appreciate the many good wishes and kind gestures.

I would like to express my thankfulness to my in-laws Ger, Diny, Jan, Marianne, Cees, Joop, Ton, John, Fanny, Marie-Anne, and Jos for your empathy, patience, kindness, understanding, and support. Dear Frans, Yvonne, and Colin, many thanks for the many advices and the candles that you lit for me.

I want to thank also Enrique, Mónica, and Moises Aguilar for your generosity and care from the distance.

Thanks a million to my nephew Manuel and niece Norah for keeping me smiley and for giving me the chance to look at life through different spectacles. I want also to acknowledge the sincere and permanent appreciatives of kindness of Tiny, Johan, and Jeroen van den Berg.

To my sisters Sylvia and Sigrid, and to my brother Martin, I would like to say *Muchísimas gracias*. Thank you very much for lending me your ears when I was upset and for teaching me ways to hold on to the end. Several times you encouraged me without words but deeds. I do appreciate a lot your understanding for the many weekends I stood away.

Finally, I must thank my dear husband. Although there were periods I was away, you managed to give me uncountable joys and smiles every morning. Without you, ending this journey would not have been possible. Your sense of humor, perseverance, patience, understanding, support, confidence, and love enabled me to complete this dissertation.

Shirley A. Ortega Azurduy, The Hague, May 7th, 2009

About the Author

Shirley Amparo Ortega Azurduy was born on October 14th, 1972 in Oruro, Bolivia. In December 1989, she finished secondary school at the German School in Oruro. From January 1990 until December 1994, she studied Industrial Engineering at the Militair School of Engineering in La Paz. She completed her studies by a thesis entitled "Study of the combined treatment of mining- and concentration plant-wastewaters". In 1998, her interest on industry-related environmental problems motivated her to do a Master on Environmental Technology at the IHE-UNESCO in Delft. From June 2001 until October 2003, she followed the postgraduate program Mathematics for Industry at the Stan Ackermans Instituut of Technische Universiteit Eindhoven. Her final project "Modelling factorial effects by using semi-parametric linear mixed models" was carried out at the Centre for Quantitative Methods (CQM) B.V in Eindhoven. From September 2004 until March 2009, Shirley was a PhD student in the group of Methodology and Statistics of the Department of the Faculty of Health, Medicine and Life Sciences at the University of Maastricht. In April 2009, she joined Centraal Bureau voor de Statistiek as statistical researcher.